

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 12:12:23 ; Search time 4552.47 Seconds
(without alignments)
17398.674 Million cell updates/sec

Title: US-09-899-718A-1
Perfect score: 3785
Sequence: 1 gttgggttcgctttttc.....tggaccggtgttcgtcgac 3785

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0
Searched: 1797656 seqs, 10463268293 residues

Word size : 0
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.in.*
- 32: em.htg.other.*
- 33: em.htgo.inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

1	3785	100.0	3785	6	AX349063	Sequence
2	658	17.4	2805	8	AB019622	Triticum
3	544	14.4	2781	8	AB023061	Triticum
4	465	12.3	2781	8	AB029063	Triticum
5	312	8.2	2186	8	TAWAXYSS	X57233 Wheat waxy
6	156	4.1	1801	8	AF113843	Triticum
7	93	2.5	2804	8	AB029064	Triticum
8	77	2.0	2028	8	AF113844	Triticum
9	77	2.0	2127	8	AF163319	Triticum
10	77	2.0	2793	8	AB029062	Triticum
11	77	2.0	2818	8	AB019623	Triticum
12	77	2.0	2834	8	AF110373	Triticum
13	72	1.9	2289	8	AF286320	Triticum
14	69	1.8	2893	8	AF110375	Aegilops
15	68	1.8	2893	6	AX349066	Sequence
16	63	1.7	1605	8	AF250137	Triticum
17	63	1.7	2886	8	AB019624	Triticum
18	60	1.6	121	6	AX325803	Sequence
19	60	1.6	121	6	AX325804	Sequence
20	59	1.6	2826	8	AF110374	Aegilops
21	52	1.4	2311	8	HWAXYR	X07932 Barley mRNA
22	52	1.4	5153	8	HWAXYG	X07931 Barley DNA
23	52	1.4	124050	8	AF474373	Hordeum v
24	47	1.2	81019	8	AB073159	Arabidops
25	44	1.2	121	6	AX325799	Sequence
26	44	1.2	121	6	AX325800	Sequence
27	41	1.1	121	6	AX325823	Sequence
28	41	1.1	121	6	AX325824	Sequence
29	39	1.0	2287	8	SHU23945	U33945 Sorghum bic
30	39	1.0	4800	6	AR106490	AR106490 Sequence
31	39	1.0	4800	8	MZEWAXY	M24258 Maize amylo
32	39	1.0	4800	8	MZEWAXY	X03935 Zea mays wa
33	39	1.0	170949	2	AC106104	AC106104 Rattus no
34	38	1.0	169048	2	AC106675	AC106675 Rattus no
35	38	1.0	215352	2	AC007305	AC007305 Mus muscu
36	37	1.0	121	6	AX325835	Sequence
37	37	1.0	121	6	AX325836	Sequence
38	37	1.0	121	6	AX325839	Sequence
39	37	1.0	121	6	AX325840	Sequence
40	37	1.0	121	6	AX325843	Sequence
41	37	1.0	121	6	AX325844	Sequence
42	37	1.0	18772	2	AC109854	AC109854 Rattus no
43	37	1.0	69729	2	AC102413	AC102413 Mus muscu
44	37	1.0	75756	9	AC068601	AC068601 Homo sapi
45	37	1.0	134472	9	AC005186	AC005186 Human Chr

ALIGNMENTS

RESULT	1	AX349063	3785 bp	DNA	linear	PAT 06-FEB-2002
AX349063	Sequence 1 from Patent WO0202785.					
LOCUS	AX349063					
DEFINITION	AX349063					
ACCESSION	AX349063.1	GI:18615098				
VERSION						
KEYWORDS	bread wheat.					
SOURCE	Triticum aestivum					
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Triticum.					
REFERENCE	1 (sites)					
AUTHORS	Sprunck,S., Kluth,A., Becker,D., Luetticke,S. and Loez,H.					
TITLE	Promoters of gene expression in plant caryopses					
JOURNAL	Patent: WO 0202785-A 1 10-JAN-2002;					
FEATURES	Aventis CropScience GmbH (DE)					
Source	Location/Qualifiers					
	1..3785					
	/organism="Triticum aestivum"					
	/db_xref="taxon:4565"					
BASE COUNT	945 a	980 c	899 g	961 t		
ORIGIN						

Query Match	100.08;	Score 3785;	DB 6;	Length 3785;
Best Local Similarity	100.08;	Prod. No. 0;		
Matches 3785;	Conservative	0;	Mismatches 0;	Indels 0; Gaps 0;

QY	1	gtttggttcgcgtgtttttcatttccttctctttaaagggtgaataccaatgacagtaat	60
DB	1	gtttgggttcgcgtgtttttcatttccttctctttaaagggtgaataccaatgacagtaat	60
QY	61	tcatttgtgaacagtcgagctctcttgccaaattatgtacaaattcttttgttaattgtt	120
DB	61	tcataatttgttaacagtcggatctcttgccaaattatgtacaaattcttttgttaattgtt	120
QY	121	tgtttcagtttttatttcatttccttctctttaaagggtgaataccaatgcccccaattca	180
DB	121	tggttttcagtttttatttcatttccttctctttaaagggtgaataccaatgcccccaattca	180
QY	181	ttctacctaaagagaaattcagtttttactactagtttcagtttttattttatttttaagt	240
DB	181	ttctacctaaagagaaattcagtttttactactagttttcagtttttattttatttttaagt	240
QY	241	gttttttagttgttttctcaatttatgtatgatgaataattagggtgtgtgctgtgt	300
DB	241	gttttttagttgttttctcaatttatgtatgatgaataattagggtgtgtgctgtgt	300
QY	301	gttaatatcacataagattatatacaccatttttgcagtcataaaattagcaatttca	360
DB	301	gtttaatatcacatagattatatacaccatttttgcagtcataaaattagcaatttca	360
QY	361	gtacaaattgtgcgcaaacctcttccttaatttttttttttttttttttttttttttaa	420
DB	361	gtacaaattgtgcgcaaacctcttccttaatttttttttttttttttttttttttttaa	420
QY	421	gggtaataccaatgataactaatttatgcctcattttggaatttcgtttgaataattatgc	480
DB	421	gggtaataccaatgataactaatttatgcctcattttggaatttcgtttgaataattatgc	480
QY	481	tagtacacacttattcttgatattatttgaaaagcgcaatttctgtgaatttttgtcat	540
DB	481	tagtacacacttattcttgatattatttgaaaagcgcaatttctgtgaatttttgtcat	540
QY	541	tcctgtatttttttctaatctttctcttggaagggtgaacactaatgcacaaattcat	600
DB	541	tcctgtatttttttctaatctttctcttggaagggtgaacactaatgcacaaattcat	600
QY	601	tcctgtctgaataaacttttagtattttgattgtgttttagttttttatttcattttgtct	660
DB	601	tcctgtctgtagaataaacttttagtattttgattgtgttttagttttttatttcattttgtct	660
QY	661	tccttaagggaataaccaatgccaactaatccattccattcttagaaaactctctttatctta	720
DB	661	tccttttaagggaataaccaatgccaactaatccattccattcttagaaaactctctttatctta	720
QY	721	caaaaactcgaacttttatatgcttattcgtgcataattataaaaagcacagttttctatctta	780
DB	721	caaaaaactcgaacttttatatgcttattcgtgcataattataaaaagcacagttttctatctta	780
QY	781	aattgcgtgcgaactttatcatattttgtctaaattattttttctagaatgatgatacc	840
DB	781	aattgcgtgcgaactttatcatattttgtctaaattattttttctagaatgatgatacc	840
QY	841	aatgccactaatcattccgtgagcagcaatatcggaatgcctacgtatatattagtgtt	900
DB	841	aatgccactaatcattcctgtgagcagcaatatcggaatgcctacgtatatattagtgtt	900
QY	901	gtcgcatttttctctcagcgaatgggcatgcataccctcacatgcacacacaagcat	960
DB	901	gtcgcattttttctctcagcgaatgggcatgcataccctcacatgcacacacagcat	960
QY	961	acacacaacatgagcactcagcgcagacatgcatcacactgtgcgcacacacagacac	1020
DB	961	acacacaacatgagcactcagcgcagacatgcatcacactgtgcgcacacacagacac	1020

QY 3459 ccg 3461
DB 383 CGG 385

RESULT 6
AFI13843 1801 bp mRNA linear PLN 20-APR-1999
LOCUS Triticum aestivum granule-bound starch synthase precursor (Wx-Al)
DEFINITION mRNA, Wx-Alb allele, complete cds.
ACCESSION AFI13843
VERSION AFI13843.1 GI:4588606
KEYWORDS .
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooidae; Triticeae; Triticum.
REFERENCE 1 bases 1 to 1801
AUTHORS Vrinten,P., Nakamura,T. and Yamamori,M.
JOURNAL Molecular characterization of waxy mutations in wheat
MEDLINE Mol. Gen. Genet. 261 (3), 463-471 (1999)
99254805
REFERENCE 2 (bases 1 to 1801)
AUTHORS Vrinten,P.L. and Nakamura,T.
TITLE Direct Submission
JOURNAL Submitted (14-DEC-1998) Crop Breeding, Tohoku National Agricultural
Experiment Station, Shimo-Kuriyagawa, Morioka, Iwate 020-01, Japan
FEATURES Location/Qualifiers
source
1..1801
/organism="Triticum aestivum"
/cultivar="Waxy"
/db_xref="taxon:4565"
gene
1..1801
/gene="Wx-Al"
/note="Non-functional Wx-Alb allele of waxy wheat;
spontaneously mutated Wx-Al gene"
allele="Wx-Alb"
32..1729
/gene="Wx-Al"
/note="Waxy protein"
CDS
/codon_start=1
/product="granule-bound starch synthase precursor"
/protein_id="AAD26155.1"
/db_xref="GI:4588607"
translation="MAALVTSQATSGTVLSTDRFRFGQLRPNPDAALIGIRTY
GVSAAPKQSRKHPRGNRCLSMVANGHRVMVISPRDYQKDAMDTSVISEIKVVDRIY
ERVFHCYRGVDVFDPHPCFLKVRGKTKEIYGPDAGTDVEDNQRFSLCCQA
LEVRLIDANNPHFSPGEDYVFVNCNDHPHTGLLACYLKSNYSNGIYRTAKVAFCI
HNISGRDFSDFDQNLNDRPKSFDFIDGTGPVEGRKLNHWKAGILQADKVLI
SPYAELISGEARGLDNIMRLTGITGVNGMDVSEWDPIKDKFLVTNYDVTLLE
GKANLKAQAEVGLPVDKRVPVIAFVIGREKQGDPMVIAAIPFIVEKDEVDVIQ
TKGKFRLLSKVEEKPTKRVAVFNAPLAHQMGADVLAVTSREFPGCLIQLOQ
MRYCTPCACATCGLDVITI VEGTGFPHMGRLSVDCNVNPEADVKKVTTLRKVVKG
TPAYHEVNKMCMODLSWKGPANWEDLVLELGVEGSEPPGIVGEIAPALENAVAA
373 a 548 c 586 g 294 t
BASE COUNT
ORIGIN

Query Match 4.1%; Score 156; DB 8; Length 1801;
Best Local Similarity 100.0%; Pred. No. 3e-74;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3109 ttgctgcagtgtagcacacccctgcgcgcgatcgcggtcctgtgtaacgtccagctgcg 3168
Db 1 TTGCTGCAGTAGCACACCTGCGCGGCCATGGCGGCTCTGGTCACGTCCAGCTCGC 60

Qy 3169 caactcggcaccgtcctcaggctacacgacagattccggcgtccagtttccaggcct 3228
Db 61 CACTCCGGACCGTCTCTCAGCGTCAACGACAGATTCCGGCGTCCAGGTTTTCCAGGCT 120

Qy 3229 gagggccccgaacccggcgatcgcggtcctggcat 3264

RESULT 8
AF113844

AUTHORS Mirai, J., Taira, T. and Ohta, D.
TITLE Isolation and characterization of the three waxy genes encoding the granule-bound starch synthase in hexaploid wheat

JOURNAL Gene 234 (1), 71-79 (1999)
MEDLINE 99321800

REFERENCE 2 (bases 1 to 2818)
AUTHORS Murai, J., Taira, T. and Ohta, D.
TITLE Direct Submission
JOURNAL Submitted (10-NOV-1998) Jun Murai, Osaka Prefecture University, College of Agriculture; 1-1 Gakuen-cho, Sakai, Osaka 599-8531, Japan (E-mail: junki@edemeter.plant.osakafu-u.ac.jp, Tel: 81-722-54-9409, Fax: 81-722-54-9409)

FEATURES
 source location/Qualifiers
 1. .2818
 /organism="Triticum aestivum"
 /db_xref="taxon:4565"
 join(13..336,436..516,605..703,817..970,1104..1204,1274..1627,1720..1899,1986..2177,2262..2348,2446..2574,2690..2806)
 /gene="waxy"
 join(13..336,436..516,605..703,817..970,1104..1204,1274..1627,1720..1899,1986..2177,2262..2348,2446..2574,2690..2806)
 /gene="waxy"
 /codon_start=1
 /product="starch synthase (GBSSI)"
 /protein_id="BAA77351.1"
 /db_xref="GI:4760582"
 /translation="MAALVTSOLATSGVILGIDRFRAGFGVGRPSADALGMRT TGASAPKQSRKAHRCRRLCSVMVYRATSGAGMNLVFGAEMAPWSKTGGDLVIG LPPAMAANGHRVWISPRYDOYKDAWDTSVSEIKVADEYERVYFHCYKRGVDVFFV DHPCFLEKVRGKTEKIYGPDAQDYEDNQLRFSLLCOAALPAEIRLDLNNPYFSGP YGDEVFVNDWHTGLLACYLKSQSGGIRYRTAKVAFCHNISYQGRFSDFAQLN LPDRFKSDFDIDGDKPVEGRKINMKAGILOAKVLTVPYAEELISGEARCEL DNIMRLTGITGVNDVSEWDPDKFLAANDVTTALEGKALKEALQAEVGLPVD RKVPLVAFTRLEQKQPDVMTAAIPEILKEEDVQIVLLGTGKKFERLLKSVEEKFP SKRVAVRFPNAPLQHMAGADVLAVTSRFEPCGLIQLOQMRGYTPCACASTGGLVDT IMEGKTGFHMRSLSDCNVVEPADVKVVTTLKRAVKVGTGPAYHEMVKMCMQIDLW KGPANWEDVLELGVGSEPGVGEETAPLAVENVAAP"

BASE COUNT 579 a 846 c 828 g 565 t
ORIGIN

Query Match 2.0%; Score 77; DB 8; Length 2818;
Best Local Similarity 100.0%; Pred. No. 2.1e-30;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3538 tgcaggccaacgggtacgggtcatgttcattcccccgcgtacaccagtagcaaggagc 3597
 |||||
 Db 431 TGCAGGCCAACGGTCACCGGGTCATGTCATCTCCCGCGCTACGACGAGTACAAGGACG 490
 |||||

QY 3598 cctgggacacacaggcgtc 3614
 |||||
 Db 491 CCTGGACACACAGCGTC 507
 |||||

RESULT 12
LOCUS AF110373
DEFINITION Triticum monococcum granule-bound starch synthase wx-Tma protein (wx-Tma) gene, complete cds.
ACCESSION AF110373
VERSION AF110373.1
KEYWORDS GI:6318537
SOURCE Triticum monococcum.
ORGANISM Triticum monococcum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Triticum.
 1 (bases 1 to 2834)
REFERENCE 1 (bases 1 to 2834)
AUTHORS Yan, L., Bhawe, M., Fairclough, R., Konic, C., Rahman, S. and Appels, R.
TITLE The genes encoding granule-bound starch synthases at the waxy loci of the A, B and D progenitors of common wheat

JOURNAL Genome (1999) In press
REFERENCE 2 (bases 1 to 2834)
AUTHORS Yan, L., Bhawe, M., Fairclough, R., Konic, C., Rahman, S. and Appels, R.
TITLE Direct Submission
JOURNAL Submitted (02-DEC-1998) School of Life Sciences and Technology, Victoria University of Technology, Werribee Campus, PO Box 14428 MCMC, Melbourne, Victoria 8001, Australia

FEATURES
 source location/Qualifiers
 1. .2834
 /organism="Triticum monococcum"
 /cultivar="AUS 22986"
 /db_xref="taxon:4568"
 join(32..355,444..524,613..711,830..983,1127..1227,1309..1662,1752..1931,2013..2204,2288..2374,2474..2602,2718..2834)
 /gene="wx-Tma"
 <32..>2834
 /gene="wx-Tma"
 /note="waxy"
 join(32..355,444..524,613..711,830..983,1127..1227,1309..1662,1752..1931,2013..2204,2288..2374,2474..2602,2718..2834)
 /gene="wx-Tma"
 /note="GBSSI"
 /codon_start=1
 /product="granule-bound starch synthase wx-Tma protein"
 /protein_id="AAF06936.1"
 /db_xref="GI:6318538"
 /translation="MAALVTSOLATSGAVILGIDRFRAGFGVGRPSADALGMRT VGASAPKQSRKAHRCRRLCSVMVYRATSGAGMNLVFGAEMAPWSKTGGDLVIG LPPAMAANGHRVWISPRYDOYKDAWDTSVSEIKVADEYERVYFHCYKRGVDVFFV DHPCFLEKVRGKTEKIYGPDAQDYEDNQLRFSLLCOAALPAEIRLDLNNPYFSGP YGDEVFVNDWHTGLLACYLKSQSGGIRYRTAKVAFCHNISYQGRFSDFAQLN LPDRFKSDFDIDGDKPVEGRKINMKAGILOAKVLTVPYAEELISGEARCEL DNIMRLTGITGVNDVSEWDPDKFLAANDVTTALEGKALKEALQAEVGLPVD RKVPLVAFTRLEQKQPDVMTAAIPEILKEEDVQIVLLGTGKKFERLLKSVEEKFP SKRVAVRFPNAPLQHMAGADVLAVTSRFEPCGLIQLOQMRGYTPCACASTGGLVDT IMEGKTGFHMRSLSDCNVVEPADVKVVTTLKRAVKVGTGPAYHEMVKMCMQIDLW KGPANWEDVLELGVGSEPGVGEETAPLAVENVAAP"

BASE COUNT 578 a 834 c 831 g 591 t
ORIGIN

Query Match 2.0%; Score 77; DB 8; Length 2834;
Best Local Similarity 100.0%; Pred. No. 2.1e-30;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3538 tgcaggccaacgggtacgggtcatgttcattcccccgcgtacaccagtagcaaggagc 3597
 |||||
 Db 439 TGCAGGCCAACGGTCACCGGGTCATGTCATCTCCCGCGCTACGACGAGTACAAGGACG 498
 |||||

QY 3598 cctgggacacacaggcgtc 3614
 |||||
 Db 499 CCTGGACACACAGCGTC 515
 |||||

RESULT 13
LOCUS AF286320
DEFINITION Triticum aestivum granule bound starch synthase I (gbssI) gene, complete cds.
ACCESSION AF286320
VERSION AF286320.1
KEYWORDS GI:11037535
SOURCE bread wheat.
ORGANISM Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Triticum.
 1 (bases 1 to 2289)
REFERENCE 1 (bases 1 to 2289)
AUTHORS McCue, K.F., Harkman, W.J., Tanaka, C.K. and Anderson, O.D.
TITLE Starch Branching Enzymes Sbe1 and Sbe2 from Wheat (Triticum aestivum cv. Cheyenne): Molecular Characterization, Developmental

Expression, and Homolog Assignment by Differential PCR	Unpublished	2 (bases 1 to 2289)	McCue,K.F. and Anderson,O.D.	Direct Submission	Submitted (11-JUL-2000)	United States Department of Agriculture, Agricultural Research Service, 800 Buchanan Street, Albany, CA 94710-1105, USA	Location/Qualifiers	1. .2289	/organism="Triticum aestivum"	/cultivar="Cheyenne"	/db_xref="taxon:4565"	/tissue_type="endosperm"	204. .2021	/gene="GbsSI"	204. .2021	/gene="GbsSI"	/EC_number="2.4.1.21"	/note="glucosyltransferase; ADP glucose glucosyl transferase; waxy; Wx"	/codon_start=1	/product="granule bound starch synthase I"	/protein_id="AAG27624.1"	/db_xref="GI:11037536"	/translation="MAALVTSQATSGTGLITDRFRAGFGQVRPSPADALGMRT TGASAPKQSKAHRTKRLCLSMVVRATGSGMNLVFGAEMAPWSTKGLGDLVGG LPMAANGHRVWISPRDQYKDAWDTSVSEIKVADYEVVRYHCYKRGVDRVFF DHPFLKVRGKTKKIKYGDAGDTEDNQLRSLCQAALAPRILNDNNPYSGP YGDVVFVNCNDHTGLLACYLKSNTQSSGIYTKAFCHINISVGGFDFDFAQLN LPRFKSSFFIDGYPKVEGRKINMKKAGILQADKLVTSYPAEELISGARGCEL DNLRLTGITGIVNGMDVSEDPADKFLAANDVDTTALGKALKEALQAEVGLPVD RYPLVAFIRGLEQKQPDVMAIAPEILKEEDVOIVLLGTGKKRPERLLKSIEKFF SKRVAVRENAPLAHOMAGADVLAVTSRFEPCGLIQLQGMRYGTPCACASTGGGLVD IMEGKTFHMGSLSDVNCVVEPADVKVVTTLKRAVKVVGTPAYHEMVKNCMIQDLSW KGPAKNEDVLELGESEPGVIGEEIAPLAMENVAAP"	480 a 684 c 735 g 390 t	BASE COUNT	584 a 684 c 735 g 390 t	ORIGIN																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
Query Match	1.9%;	Score 72;	DB 8;	Length 2289;																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																		

BASE COUNT 15 a 20 c 17 g 20 t
ORIGIN

Query Match 1.8%; Score 68; DB 6; Length 72;
Best Local Similarity 100.0%; Pred. No. 1.6e-25;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1936 cctttcaggacgatgcttcggtgccttaagacacacacctacacctttgtgtctatgacatgtgag 1995
Db 1 CCTTTCAGGACGATGCTTCGGTGCCTTAAGACACCTACCTTTGTGTCTATGACATGTGAG 60
Qy 1996 cccaacag 2003
Db 61 CCCAACAG 68

Search completed: July 31, 2002, 16:11:39
Job time: 14356 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 12:26:33 ; Search time 79.86 Seconds
(without alignments)
11641.908 Million cell updates/sec

Title: US-09-899-718A-1
Perfect score: 3785
Sequence: 1 gtttggttcgctgttttc.....tgaccgcgtgttcgtcgac 3785

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 383533 seqs, 122816752 residues

Word size : 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	1.0	2267	4	US-08-679-645-25
2	39	1.0	4800	3	US-08-941-445A-4
3	31	0.8	43795	3	US-08-742-185-101
4	25	0.7	252	2	US-08-623-906A-1
5	25	0.7	2542	3	US-08-941-445A-6
6	24	0.6	39	4	US-08-860-038-23
7	24	0.6	39	4	US-09-580-923-23
8	24	0.6	1865	4	US-09-370-253-5
9	24	0.6	3211	2	US-08-574-959A-8
10	24	0.6	3211	2	US-09-357-014-8
11	24	0.6	3901	2	US-08-574-959A-6
12	24	0.6	3901	4	US-09-357-014-6
13	24	0.6	5394	3	US-08-688-376-1
14	23	0.6	2621	2	US-08-553-619B-8
15	23	0.6	3387	1	US-08-261-822A-5
16	23	0.6	3387	5	PCT-US95-07744A-5
17	22	0.6	25	3	US-08-894-511-5
18	22	0.6	25	4	US-08-860-038-1
19	22	0.6	25	4	US-08-860-038-12
20	22	0.6	25	4	US-09-580-923-1
21	22	0.6	36	3	US-08-941-445A-29
22	22	0.6	51	4	US-09-580-923-33
23	22	0.6	57	3	US-08-894-511-3
24	22	0.6	57	3	US-08-894-511-4
25	22	0.6	57	3	US-08-894-511-18
26	22	0.6	58	4	US-08-860-038-9
27	22	0.6	58	4	US-08-860-038-10

28	22	0.6	58	4	US-09-580-923-9	Sequence 9, Appli
29	22	0.6	58	4	US-09-580-923-10	Sequence 10, Appl
30	22	0.6	62	6	5457089-7	Patent No. 5457089
31	22	0.6	66	3	US-08-894-511-16	Sequence 16, Appl
32	22	0.6	66	3	US-08-894-511-17	Sequence 17, Appl
33	22	0.6	66	4	US-08-860-038-20	Sequence 20, Appl
34	22	0.6	66	4	US-08-860-038-21	Sequence 21, Appl
35	22	0.6	66	4	US-09-580-923-20	Sequence 20, Appl
36	22	0.6	66	4	US-09-580-923-21	Sequence 21, Appl
37	22	0.6	188	1	US-08-115-497-21	Sequence 21, Appl
38	22	0.6	188	1	US-08-466-670-21	Sequence 21, Appl
39	22	0.6	188	2	US-08-291-011-1	Sequence 21, Appl
40	22	0.6	188	4	US-09-266-065-1	Sequence 1, Appli
41	22	0.6	405	2	US-08-299-074A-1	Sequence 1, Appli
42	22	0.6	405	4	US-09-399-773-1	Sequence 1, Appli
43	22	0.6	550	4	US-09-306-042-2	Sequence 2, Appli
44	22	0.6	678	4	US-09-227-357-81	Sequence 81, Appl
45	22	0.6	884	1	US-08-365-981-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-08-679-645-25
; Sequence 25, Application US/08679645
; Patent No. 6350934
; GENERAL INFORMATION:
; APPLICANT: Zwick, Michael G.
; APPLICANT: Edington, Brent E.
; APPLICANT: McSwiggen, James A.
; APPLICANT: Merlo, Patricia Ann Owens
; APPLICANT: Guo, Lining
; APPLICANT: Skokut, Thomas A.
; APPLICANT: Young, Scott A.
; APPLICANT: Folkerts, Otto
; APPLICANT: Merlo, Donald J.
; TITLE OF INVENTION: COMPOSITION AND METHODS FOR
; TITLE OF INVENTION: MODULATION OF GENE EXPRESSION
; TITLE OF INVENTION: IN PLANTS
; NUMBER OF SEQUENCES: 1263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/679,645
; FILING DATE: July 12, 1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/001,135
; FILING DATE: July 13, 1995
; APPLICATION NUMBER: 08/300,726
; FILING DATE: September 2, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 219/247
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 25:

Thu Aug 1 08:29:55 2002

Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3576 cgctacgaccagtagcagcctgggacacacgagcgc 3614
|||||
Db 1718 CGCTACGACCAGTACAGGACGCTGGGACACACGAGC 1756
|||||

RESULT 3

US-08-742-185-101
; Sequence 101, Application US/08742185
; Patent No. 6020476
; GENERAL INFORMATION:
; APPLICANT: Page, David C.
; APPLICANT: Reijo, Renee
; APPLICANT: Saxena, Richa
; APPLICANT: Hawkins, Trevor
; APPLICANT: Reeve, Mary Pat
; TITLE OF INVENTION: DAZ: A GENE FAMILY ASSOCIATED WITH AZOOSPERMIA
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: US
; ZIP: 02173

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/742,185
FILING DATE: 30-OCT-1996
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER: US 08/690,734
FILING DATE: 31-JUL-1996
APPLICATION DATA:
APPLICATION NUMBER: US 08/310,429
FILING DATE: 22-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: WHI94-07A2
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 101:
SEQUENCE CHARACTERISTICS:
LENGTH: 43795 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-742-185-101

Query Match 0.8%; Score 31; DB 3; Length 43795;
Best Local Similarity 100.0%; Pred. No. 4.4e-05;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2385 gaagaagaagaagaagaagaagaagaagaag 2415
|||||
Db 38989 GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 39019
|||||

RESULT 4

US-08-623-906A-1/C
; Sequence 1, Application US/08623906A
; Patent No. 5874217
; GENERAL INFORMATION:
; APPLICANT: Stevenson, Tamara

SEQUENCE CHARACTERISTICS:
LENGTH: 2267 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-679-645-25

Query Match 1.0%; Score 39; DB 4; Length 2267;
Best Local Similarity 100.0%; Pred. No. 6.8e-09;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3576 cgctacacagtagcagcctgggacacacgagcgc 3614
|||||
Db 543 CGCTACGACCAGTACAGGACGCTGGGACACACGAGC 581
|||||

RESULT 2

US-08-941-445A-4
; Sequence 4, Application US/08941445A
; Patent No. 6107060
; GENERAL INFORMATION:
; APPLICANT: Keeling, Peter
; APPLICANT: Guan, Haining
; TITLE OF INVENTION: Starch Encapsulation
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle
; CITY: Boulder
; STATE: CO
; COUNTRY: US
; ZIP: 80303

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/941,445A
FILING DATE: 30-SEP-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,855
FILING DATE: 30-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Winner, Ellen P
REGISTRATION NUMBER: 28,547
REFERENCE/DOCKET NUMBER: 89-97
TELEPHONE: (303) 499-8080
TELEFAX: (303) 499-8089
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 4800 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: not relevant
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Zea mays

FEATURE:
NAME/KEY: CDS
LOCATION: Join(1449..1553, 1685..1765, 1860..1958, 2055
..2144, 2226..2289, 2413..2513, 2651..2760, 2858
..3101, 3212..3394, 3490..3681, 3793..3879, 3977
..4105, 4227..4343)
LOCATION: ..4105, 4227..4343)
US-08-941-445A-4

Query Match 1.0%; Score 39; DB 3; Length 4800;
Best Local Similarity 100.0%; Pred. No. 6.9e-09;

Thu Aug 1 08:29:55 2002

us-09-899-718a-1.oli.rni

Db 8 aggaagaagaagaagaagaag 31

RESULT 8

US-09-370-253-5/c

Sequence 5, Application US/09370253

Patent No. 6165792

GENERAL INFORMATION:

APPLICANT: Allen, Stephen M.

APPLICANT: Sakai, Hajime

APPLICANT: Thorpe, Catherine J.

TITLE OF INVENTION: Amino Acid Transporters

FILE REFERENCE: BB-1200

CURRENT APPLICATION NUMBER: US/09/370,253

CURRENT FILING DATE: 1999-08-09

EARLIER APPLICATION NUMBER: 60/097,222

EARLIER FILING DATE: August 20, 1998

NUMBER OF SEQ ID NOS: 18

SOFTWARE: Microsoft Office 97

SEQ ID NO 5

LENGTH: 1865

TYPE: DNA

ORGANISM: Oryza sativa

US-09-370-253-5

Query Match 0.6%; Score 24; DB 4; Length 1865;

Best Local Similarity 100.0%; Pred. No. 0.09; 0; Indels 0; Gaps 0;

Matches 24; Conservative 0; Mismatches 0;

QY 2377 aggaagaagaagaagaagaag 2400

|||||

Db 50 AGGAAGAAGAAGAAGAAGAAG 27

RESULT 9

US-08-574-959A-8

Sequence 8, Application US/08574959A

Patent No. 5962224

GENERAL INFORMATION:

APPLICANT: Jaekyoon Shin, Insil Joung, Ratna K. Vadlamudi

APPLICANT: and Jack L. Strominger

TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES

TITLE OF INVENTION: AND USES THEREFOR

NUMBER OF SEQUENCES: 22

CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD

STREET: 60 State Street, Suite 510

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/574,959A

FILING DATE: 19-DEC-95

ATTORNEY/AGENT INFORMATION:

NAME: Mandragouras, Amy E.

REGISTRATION NUMBER: 36,207

REFERENCE/DOCKET NUMBER: DFN-008

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)227-7400

TELEFAX: (617)227-5941

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 3211 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

APPLICATION NUMBER: US/08/860,038

FILING DATE: 08-NOV-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

FR 94/15162

FILING DATE: 16-DEC-1994

PRIOR APPLICATION DATA:

WO FR95/01468

FILING DATE: 08-NOV-1995

ATTORNEY/AGENT INFORMATION:

NAME: Savitzky Esq., Martin F.

REGISTRATION NUMBER: 29,699

REFERENCE/DOCKET NUMBER: ST94090-US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (610) 454-3816

TELEFAX: (610) 454-3808

INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:

LENGTH: 39 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid

DESCRIPTION: /desc = "Oligonucleotide"

US-08-860-038-23

Query Match 0.6%; Score 24; DB 4; Length 39;

Best Local Similarity 100.0%; Pred. No. 0.089; 0; Indels 0; Gaps 0;

Matches 24; Conservative 0; Mismatches 0;

QY 2377 aggaagaagaagaagaagaag 2400

|||||

Db 8 AGGAAGAAGAAGAAGAAGAAG 31

RESULT 7

US-09-580-923-23

Sequence 23, Application US/09580923

Patent No. 6319672

GENERAL INFORMATION:

APPLICANT: Crouzet, Joel

APPLICANT: Scherman, Daniel

APPLICANT: Wills, Pierre

APPLICANT: Cameron, Beatrice

APPLICANT: Blanche, Francis

TITLE OF INVENTION: PURIFICATION OF A TRIPLE HELIX FORMATION WITH AN

TITLE OF INVENTION: IMMOBILIZED OLIGONUCLEOTIDE

FILE REFERENCE: 03804.0138-01

CURRENT APPLICATION NUMBER: US/09/580,923

CURRENT FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 08/860,038

PRIOR FILING DATE: 1997-06-09

PRIOR APPLICATION NUMBER: PCT/FR95/01468

PRIOR FILING DATE: 1995-11-08

NUMBER OF SEQ ID NOS: 36

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 23

LENGTH: 39

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence:

OTHER INFORMATION: oligonucleotide

US-09-580-923-23

Query Match 0.6%; Score 24; DB 4; Length 39;

Best Local Similarity 100.0%; Pred. No. 0.089; 0; Indels 0; Gaps 0;

Matches 24; Conservative 0; Mismatches 0;

QY 2377 aggaagaagaagaagaagaag 2400

|||||

; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 439..3157
US-08-574-959A-8

Query Match 0.6%; Score 24; DB 2; Length 3211;
Best Local Similarity 100.0%; Pred. No. 0.09;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2377 aggaagaagaagaagaagaag 2400
|||||
DB 2423 AGGAAGAAGAAGAAGAAG 2446

RESULT 10
US-09-357-014-8
; Sequence 8, Application US/09357014
; Patent No. 6291645
; GENERAL INFORMATION:
; APPLICANT: Jaekyoon Shin, Insil Joung, Ratna K. Vadlamudi
; and Jack L. Strominger
; TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES
; AND USES THEREFOR
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/357,014
; FILING DATE: 19-Jul-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/574,959
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: DFN-008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3211 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 439..3157
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:

Query Match 0.6%; Score 24; DB 4; Length 3211;
Best Local Similarity 100.0%; Pred. No. 0.09;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2377 aggaagaagaagaagaagaag 2400
|||||
DB 2423 AGGAAGAAGAAGAAGAAG 2446

RESULT 11
US-08-574-959A-6
; Sequence 6, Application US/08574959A
; Patent No. 5962224
; GENERAL INFORMATION:
; APPLICANT: Jaekyoon Shin, Insil Joung, Ratna K. Vadlamudi
; and Jack L. Strominger
; TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES
; AND USES THEREFOR
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/574,959A
; FILING DATE: 19-DEC-95
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: DFN-008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3901 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 439..3847
US-08-574-959A-6

Query Match 0.6%; Score 24; DB 2; Length 3901;
Best Local Similarity 100.0%; Pred. No. 0.091;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2377 aggaagaagaagaagaagaag 2400
|||||
DB 3113 AGGAAGAAGAAGAAGAAG 3136

RESULT 12
US-09-357-014-6
; Sequence 6, Application US/09357014
; Patent No. 6291645
; GENERAL INFORMATION:
; APPLICANT: Jaekyoon Shin, Insil Joung, Ratna K. Vadlamudi
; and Jack L. Strominger
; TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES
; AND USES THEREFOR
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION NUMBER: US/09/357,014
;; FILING DATE: 19-Jul-1999
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/574,959
;; FILING DATE: <Unknown>
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Mandragouras, Amy E.
;; REGISTRATION NUMBER: 36,207
;; REFERENCE/DOCKET NUMBER: DFN-008
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (617)227-7400
;; TELEFAX: (617)227-5941
;; INFORMATION FOR SEQ ID NO: 6:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 3901 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; MOLECULE TYPE: cDNA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 439..3847
;; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-357-014-6

Query Match 0.6%; Score 24; DB 4; Length 3901;
Best Local Similarity 100.0%; Pred. No. 0.091;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2377 aggaagaagaagaagaagaag 2400
|||||
DB 3113 AGGAAGAAGAAGAAGAAG 3136

RESULT 13
US-08-688-376-1
;; Sequence 1, Application US/08688376
;; Patent No. 6018039
;; GENERAL INFORMATION:
;; APPLICANT: Satow, Hiroyasu
;; TITLE OF INVENTION: NOVEL PROCESS FOR PRODUCING SUBSTANCES
;; TITLE OF INVENTION: IN MAMMARY GLAND OF TRANSGENIC ANIMAL BY USING MC26 GENE
;; NUMBER OF SEQUENCES: 8
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: DILWORTH & BARRESE
;; STREET: 4350 LaJolla Village Drive, Suite 300
;; CITY: San Diego
;; STATE: CA
;; COUNTRY: USA
;; ZIP: 92122
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/688,376
;; FILING DATE: 30-JUL-1996
;; CLASSIFICATION: 800
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Pepper, Frederick W.
;; REGISTRATION NUMBER: 31,286
;; REFERENCE/DOCKET NUMBER: 567-3
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 619-546-4410
;; TELEFAX: 619-453-2839

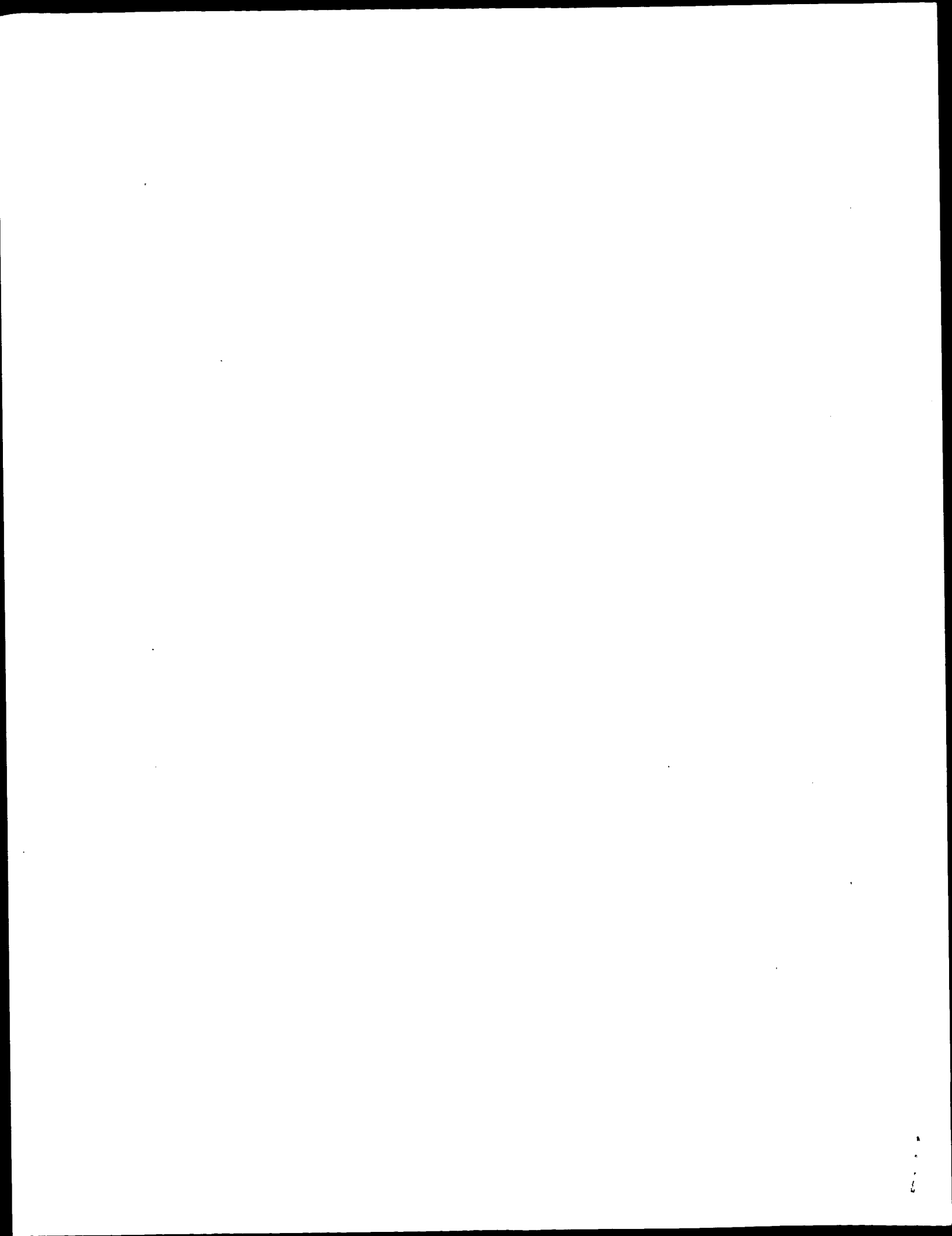
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 5394 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; FEATURE:
;; NAME/KEY: CAAT_signal
;; LOCATION: 2234..2243
;; FEATURE:
;; NAME/KEY: TATA_signal
;; LOCATION: 2275..2281
;; FEATURE:
;; NAME/KEY: polyA_signal
;; LOCATION: 4607..4612
;; US-08-688-376-1

Query Match 0.6%; Score 24; DB 3; Length 5394;
Best Local Similarity 100.0%; Pred. No. 0.091;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2377 aggaagaagaagaagaagaag 2400
|||||
DB 1437 AGGAAGAAGAAGAAGAAG 1460

RESULT 14
US-08-553-619B-8/C
;; Sequence 8, Application US/08553619B
;; Patent No. 5919705
;; GENERAL INFORMATION:
;; APPLICANT: Dehaan, Petrus T.
;; TITLE OF INVENTION: Virus Resistant Plants
;; NUMBER OF SEQUENCES: 30
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: No. 5919705artis Crop Protection
;; STREET: 975 California Avenue
;; CITY: Palo Alto
;; STATE: CA
;; COUNTRY: USA
;; ZIP: 94304
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/553,619B
;; FILING DATE: December 1, 1995
;; CLASSIFICATION: 800
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Marcus-Wyner, Lynn
;; REGISTRATION NUMBER: 34,869
;; REFERENCE/DOCKET NUMBER: 137-1082/PCT
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 415/354-3588
;; TELEFAX: 415/857-1125
;; INFORMATION FOR SEQ ID NO: 8:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2621 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: unknown
;; ORIGINAL SOURCE:
;; ORGANISM: Chimeric tomato spotted wilt virus S RNA
;; US-08-553-619B-8

Query Match 0.6%; Score 23; DB 2; Length 2621;
Best Local Similarity 100.0%; Pred. No. 0.27;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



Query Match 1.4%; Score 52; DB 9; Length 462;
Best Local Similarity 100.0%; Pred. No. 3.4e-13;

ORGANISM Hordeum vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
 ; Triticeae; Hordeum.
 REFERENCE 1 (bases 1 to 558)
 AUTHORS Wing,R., Close,T.J., Kleinbolls,A., Wise,R., Begum,D., Frisch,D., Yu
 ,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton
 ,R.D., Close,S.J., Oates,R. and Main,D.
 TITLE Development of a genetically and physically anchored EST resource
 for barley genomics: Morex 5-45 DAP spike cDNA library
 JOURNAL Unpublished (2001)
 COMMENT On Aug 21, 2000 this sequence version replaced gi:9859804.
 Contact: Wing RA
 Clemson University
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 Total hq bases = 215
 Seq primer: AATTAACTCTACTAAAGGG
 High quality sequence start: 16
 High quality sequence stop: 550.

FEATURES
 source
 1..558
 Location/Qualifiers
 /organism="Hordeum vulgare"
 /cultivar="Morex"
 /db_xref="taxon:4513"
 /clone="HVSHEH0098D06f"
 /cclone_lib="Hordeum vulgare 5-45 DAP spike EST library
 HVCdNA0009 (5 to 45 DAP)"
 /tissue_type="5-45 DAP Spike"
 /lab_host="SOLR"
 /note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI;
 Plants were grown in the greenhouse at the University of
 California, Riverside (Fenton, SJ Close, TJ Close). Whole
 spikes with awns trimmed were collected at 5, 10, 15, 20,
 30 and 45 DAP (Fenton). Total RNA was prepared from each
 pool, equal quantities of all six RNA pools were combined,
 poly(A) RNA was purified from the mixture, one primary
 unamplified cDNA library was made, and 1 million pfu were
 in vivo excised to give pBluescript SK(-) cDNA phagemids
 (Choi) in the TJ Close lab at the University of California,
 Riverside. Phagemids were plated and picked at the Clemson
 University Genomics Institute (CUGI) (Begum, Palmer,
 Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
 sequencing and sequence analysis were performed at CUGI
 (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).
 The sequence has been trimmed to remove vector sequence
 and contains a minimum of 100 bases of phred value 20 or
 above. For more details on library preparation and
 sequence analysis see
 http://www.genome.clemson.edu/projects/barley. To order
 this clone see http://www.genome.clemson.edu/orders/Also
 see close TJ, Wing R, Kleinbolls A, Wise R (2001)
 Genetically and physically anchored EST resources for
 barley genomics. Barley Genetics Newsletter 31:29-30.
 (http://wheat.pw.usda.gov/gvpages/bgn/31/cover.html)".

BASE COUNT 102 a 209 c 188 g 78 t 1 others
 ORIGIN
 Query Match 1.4%; Score 52; DB 10; Length 558;
 Best Local Similarity 100.0%; Pred. No. 3.3e-13;
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3363 gcggcatgaacctgtgttcgcccgcgagatggccctggagcaagac 3414
 |||||
 Db 440 GCGCATGAACCTCGTTCGTTCGCGCGGAGATGGCGCCTGGAGCAAGAC 491

RESULT 5
 AL508959

LOCUS AL508959 638 bp mRNA linear EST 04-JAN-2001
 DEFINITION AL508959 Hordeum vulgare Barke developing caryopsis (3.-15.DAP)
 Hordeum vulgare cDNA clone HY10E20V 5', mRNA sequence.
 ACCSSION AL508959
 VERSION AL508959.1 GI:12035462
 KEYWORDS EST.
 SOURCE barley.
 ORGANISM Hordeum vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
 ; Triticeae; Hordeum.
 REFERENCE 1 (bases 1 to 638)
 AUTHORS Michalek,W., Weschke,W., Pleissner,K.-P. and Graner,A.
 TITLE EST sequencing and analysis in barley
 JOURNAL Unpublished (2000)
 COMMENT Contact: Michalek W
 Institute for Plant Genetics and Crop Plant Research
 Corrensstr.3, D-06466 Gatersleben, Germany
 Email: michalek@ipk-gatersleben.de, http://pgrc.ipk-gatersleben.de
 Seq primer: T3 primer for 5'end.

FEATURES
 source
 1..638
 Location/Qualifiers
 /organism="Hordeum vulgare"
 /cultivar="Barke"
 /db_xref="taxon:4513"
 /clone="HY10E20V"
 /cclone_lib="Hordeum vulgare Barke developing caryopsis
 (3.-15.DAP)"
 /tissue_type="developing caryopsis (3.-15.DAP)"
 /lab_host="XLOLR"
 /note="Vector: plasmid pBK-CMV; Site_1: EcoRI; Site_2:
 XhoI; mRNA was made from developing caryopsis (3.-15.DAP)
 of spring barley variety 'Barke', a high quality malting
 variety. Cloning sites: EcoRI (5'-end of cDNA) and XhoI
 (3'-end of cDNA). NOTE: Due to a cloning artefact caused
 by the kit, in most cases the EcoRI site is NOT present,
 as well as the EcoRI adapter. Average insert size is 1 kb
 sequence trimming: Vector sequences and sequence ends were
 trimmed from the 5'-and 3'-end until a 50 bp window
 contains less than two ambiguities. The maximum length was
 set to 700 bp".

BASE COUNT 119 a 235 c 190 g 93 t 1 others
 ORIGIN
 Query Match 1.4%; Score 52; DB 9; Length 638;
 Best Local Similarity 100.0%; Pred. No. 3.2e-13;
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3363 gcggcatgaacctgtgttcgcccgcgagatggccctggagcaagac 3414
 |||||
 Db 447 GCGCATGAACCTCGTTCGTTCGCGCGGAGATGGCGCCTGGAGCAAGAC 498

RESULT 6
 AL506567 700 bp mRNA linear EST 04-JAN-2001
 LOCUS AL506567 Hordeum vulgare Barke developing caryopsis (3.-15.DAP)
 DEFINITION Hordeum vulgare cDNA clone HY03121T 5', mRNA sequence.
 ACCSSION AL506567
 VERSION AL506567.1 GI:12032782
 KEYWORDS EST.
 SOURCE barley.
 ORGANISM Hordeum vulgare

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
 ; Triticeae; Hordeum.
 REFERENCE 1 (bases 1 to 700)
 AUTHORS Michalek,W., Weschke,W., Pleissner,K.-P. and Graner,A.
 TITLE EST sequencing and analysis in barley
 JOURNAL Unpublished (2000)
 COMMENT Contact: Michalek W
 Institute for Plant Genetics and Crop Plant Research

source

1. .494
 /organism="Triticum aestivum"
 /cultivar="Wyuna"
 /db_xref="taxon:4565"
 /clone="P51-2A"
 /clone_lib="BRY"
 /cell_type="endosperm"
 98 a 170 c 160 g 66 t

BASE COUNT
 ORIGIN

Query Match 1.1%; Score 41; DB 9; Length 494;
 Best Local Similarity 100.0%; Pred. No. 5.5e-08;
 Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3322 ccggcggtgctctccatgtgtgtgcgcacacggcagcg 3362
 |||
 Db 391 CCGGCGGTGCTCTCCATGTGTGTGCGCCGCGGCGGCGC 431

RESULT 12

BM368238

LOCUS

DEFINITION EBed01_SQ002_J02_R IGF Barley EBed01 library Hordeum vulgare cDNA
 clone EBed01_SQ002_J02 5', mRNA sequence.

ACCESSION

BM368238

VERSION

EST.

KEYWORDS

SOURCE

ORGANISM

Hordeum vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
 ; Triticeae; Hordeum.

REFERENCE

1 (bases 1 to 444)

AUTHORS

Hedley, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardle, L.,
 Ramsay, L., Machray, G., Marshall, D.F.M. and Waugh, R.

TITLE

Development of Barley Transcriptome Resources

JOURNAL

Unpublished (2001)

COMMENT

Contact: Waugh R

Scottish Crop Research Institute

Invergowrie, Dundee, DD2 5DA, Scotland, UK

Tel: 00 44 1382 562731

Fax: 00 44 1382 562426

Email: rwaugh@scri.sari.ac.uk

All sequence has a Phred quality score of 20 or over

Seq primer: M13 reverse.

FEATURES

source

Location/Qualifiers

1. .444

/organism="Hordeum vulgare"

/cultivar="Optic"

/db_xref="taxon:4513"

/clone="EBed01_SQ002_J02"

/clone_lib="IGF Barley EBed01 library"

/tissue_type="Endosperm"

/dev_stage="6 days post anthesis"

/lab_host="PH108"

/note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;
 Non-normalised library, directionally cloned into pSPORT1.
 Derived from endosperm tissue dissected from developing
 grains (6 days post anthesis) in glasshouse grown barley
 plants. Developed as part of the barley transcriptome
 resources of BSRC/SEERAD funded cereal IGF (Investigating
 Gene Function) project."

BASE COUNT 90 a 160 c 131 g 63 t

ORIGIN

Query Match 1.1%; Score 40; DB 10; Length 444;

Best Local Similarity 100.0%; Pred. No. 1.7e-07;

Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3189 ggcgcacgacagattccggcggttttcaggcct 3228

Db 244

CGGTCCGCGACAGATTCGCGCGTCCAGGTTTTCAGGCGCT 283

RESULT 13

AZ733073/c

LOCUS

DEFINITION

RPCI-24-118F9-TV RPCI-24 Mus musculus genomic clone RPCI-24-118F9,
 DNA sequence.

ACCESSION

AZ733073

VERSION

GSS.

KEYWORDS

SOURCE

ORGANISM

Mus musculus

REFERENCE

1 (bases 1 to 165)

AUTHORS

Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akinret, B., Levins, M.,
 Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregorgis, E.,
 Russell, D., de Jong, P. and Fraser, C.M.

TITLE

Mouse BAC End Sequences from Library RPCI-24

JOURNAL

Unpublished (1999)

COMMENT

Other_GSSs: RPCI-24-118F9.TJ

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC
 library availability, please contact Pieter de Jong
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC
 Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
 page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 118 Row: F Column: 9
 Seq primer: T7
 Class: BAC ends.

FEATURES

source

Location/Qualifiers

1. .165

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-24-118F9"

/clone_lib="RPCI-24"

/sex="Male"

/cell_type="Spleen/Brain"

/note="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI;
 RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
 library was cloned in the pTARBAC1 cloning vector at the
 BamHI sites using MboI partially digested male C57BL/6J
 DNA."

BASE COUNT 19 a 44 c 12 g 90 t

ORIGIN

Query Match 1.0%; Score 37; DB 12; Length 165;

Best Local Similarity 100.0%; Pred. No. 5.3e-06;

Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2379 gaagaagaagaagaagaagaagaagaagaagaagaag 2415
 |||

Db 146

GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 110

RESULT 14

AZ462781/c

LOCUS

DEFINITION

1M0271G17F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 clone UUGC1M0271G17 F, DNA sequence.

ACCESSION

AZ462781

VERSION

GSS.

KEYWORDS

AZ462781.1 GI:10620906

```

SOURCE      house mouse.
ORGANISM    Mus musculus
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
              1 (bases 1 to 260)
              Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
              Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly
              M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
              and Wright, D., Weiss, R.
TITLE        Mouse whole genome scaffolding with paired end reads from 10kb
              plasmid inserts
JOURNAL      Unpublished (2000)
COMMENT      Contact: Robert B. Weiss
              University of Utah Genome Center
              University of Utah
              Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
              84112, USA
              Tel: 801 585 5606
              Fax: 801 585 7177
              Email: ddunn@genetics.utah.edu
              Insert Length: 10000 Std Error: 0.00
              Plate: 0271 row: G column: 17
              Seq primer: CGTTGTAACGACGCCAGT
              Class: plasmid ends
              High quality sequence stop: 260.
              Location/Qualifiers
FEATURES     source
              1..260
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="UUGC1M0271G17"
                /clone.lib="Mouse 10kb plasmid UUGC1M library"
                /sex="Male"
                /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                /note="Vector: PWD42nv; Purified genomic DNA from M.
              musculus C57BL/6J (male) was obtained from the Jackson
              Laboratory Mouse DNA Resource
              (http://www.jax.org/resources/documents/dnares/). The DNA
              was hydrodynamically sheared by repeated passage through a
              0.005 inch orifice at constant velocity. The sheared DNA
              was blunt end-repaired with T4 DNA polymerase and T4
              polynucleotide kinase. Adaptor oligonucleotides were
              ligated to the blunt ends in high molar excess. The
              adapted DNA was purified and size-selected for a 9.5 to
              10.5 kb range using preparative agarose gel
              electrophoresis. Vector DNA was prepared from a derivative
              of pWD42 (gi14732114|gb|AF129072.1), a copy-number
              inducible derivative of plasmid R1. The vector was ligated
              with adaptors complementary to the insert adaptors and
              purified. The sheared, adapted mouse DNA was annealed to
              adapted vector DNA, and transformed into
              chemically-competent E. coli XL10-Gold (Stratagene) cells
              and selected for ampicillin resistance."
BASE COUNT   49 a 75 c 30 g 106 t
ORIGIN
Query Match      1.0%; Score 37; DB 12; Length 260;
Best Local Similarity 100.0%; Pred. No. 4.9e-06;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2379 gaagaagaagaagaagaagaagaagaagaagaagaag 2415
      |||||||
Db 180 GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 144
      |||||||
RESULT 15
AZ368386/c
LOCUS      380 bp DNA linear GSS 02-OCT-2000
DEFINITION lM0118M19F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0118M19 F, DNA sequence.
ACCESSION AZ368386
VERSION    AZ368386.1 GI:10482086

```

```

KEYWORDS     GSS.
SOURCE        house mouse.
ORGANISM      Mus musculus
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
              1 (bases 1 to 380)
              Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
              Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly
              M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
              and Wright, D., Weiss, R.
TITLE          Mouse whole genome scaffolding with paired end reads from 10kb
              plasmid inserts
JOURNAL        Unpublished (2000)
COMMENT        Contact: Robert B. Weiss
              University of Utah Genome Center
              University of Utah
              Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
              84112, USA
              Tel: 801 585 5606
              Fax: 801 585 7177
              Email: ddunn@genetics.utah.edu
              Insert Length: 10000 Std Error: 0.00
              Plate: 0118 row: M column: 19
              Seq primer: CGTTGTAACGACGCCAGT
              Class: plasmid ends
              High quality sequence stop: 380.
              Location/Qualifiers
FEATURES       source
              1..380
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="UUGC1M0118M19"
                /clone.lib="Mouse 10kb plasmid UUGC1M library"
                /sex="Male"
                /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                /note="Vector: PWD42nv; Purified genomic DNA from M.
              musculus C57BL/6J (male) was obtained from the Jackson
              Laboratory Mouse DNA Resource
              (http://www.jax.org/resources/documents/dnares/). The DNA
              was hydrodynamically sheared by repeated passage through a
              0.005 inch orifice at constant velocity. The sheared DNA
              was blunt end-repaired with T4 DNA polymerase and T4
              polynucleotide kinase. Adaptor oligonucleotides were
              ligated to the blunt ends in high molar excess. The
              adapted DNA was purified and size-selected for a 9.5 to
              10.5 kb range using preparative agarose gel
              electrophoresis. Vector DNA was prepared from a derivative
              of pWD42 (gi14732114|gb|AF129072.1), a copy-number
              inducible derivative of plasmid R1. The vector was ligated
              with adaptors complementary to the insert adaptors and
              purified. The sheared, adapted mouse DNA was annealed to
              adapted vector DNA, and transformed into
              chemically-competent E. coli XL10-Gold (Stratagene) cells
              and selected for ampicillin resistance."
BASE COUNT     51 a 116 c 18 g 195 t
ORIGIN
Query Match      1.0%; Score 37; DB 12; Length 380;
Best Local Similarity 100.0%; Pred. No. 4.5e-06;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2379 gaagaagaagaagaagaagaagaagaagaagaagaag 2415
      |||||||
Db 373 GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 337
      |||||||
Search completed: July 31, 2002, 14:54:43
Job time: 15085 sec

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 14:02:43 ; Search time 389.91 Seconds
(without alignments)
16666.721 Million cell updates/sec

Title: US-09-899-718A-1
Perfect score: 3785
Sequence: 1 gtttggttcgtgttttc.....tgaccggtgttcgtgcac 3785

Scoring table:
OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_032802.*

- 1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	1.0	333	21 AAC74569	Human ORFX ORF124
2	39	1.0	1915	19 AAX60319	DNA sequence of th
3	39	1.0	2267	18 AAX63355	Granule bound star
4	39	1.0	4800	19 AAV29752	Zea mays waxy gene
C 5	37	1.0	276	22 ABA68266	Human foetal liver
C 6	37	1.0	276	22 ABA35275	Probe #13741 for g
C 7	37	1.0	276	22 AAK42405	Human bone marrow
C 8	37	1.0	276	22 AAI23164	Probe #13097 for g
9	37	1.0	460	23 AAS71148	DNA encoding novel

10	37	1.0	508	23 AAS69548	DNA encoding novel
11	37	1.0	771	23 AAS90697	DNA encoding novel
12	37	1.0	2187	23 AAS75468	DNA encoding novel
13	31	0.8	711	23 AAS90716	DNA encoding novel
14	31	0.8	43795	21 AAZ92583	Human DAZ genomic
15	29	0.8	509	21 AAC292802	Arabidopsis thalia
16	29	0.8	512	21 AAC39876	Arabidopsis thalia
17	29	0.8	571	21 AAC52152	Arabidopsis thalia
18	29	0.8	825	21 AAC55083	Arabidopsis thalia
19	28	0.7	621	23 AAS90688	DNA encoding novel
20	27	0.7	299	23 AAS71084	DNA encoding novel
21	27	0.7	435	23 AAS67160	DNA encoding novel
22	27	0.7	435	23 AAS68115	DNA encoding novel
23	27	0.7	435	23 AAS69544	DNA encoding novel
24	27	0.7	435	23 AAS70705	DNA encoding novel
25	27	0.7	435	23 AAS71149	DNA encoding novel
26	26	0.7	459	21 AAC52482	Arabidopsis thalia
27	26	0.7	988	21 AAC56084	Eucalyptus grandis
28	26	0.7	1083	21 AAC44171	Arabidopsis thalia
29	26	0.7	1227	21 AAC40176	Arabidopsis thalia
30	26	0.7	86584	21 AAF22292	BAC containing rep
31	25	0.7	140	21 AAC13041	Human secreted pro
32	25	0.7	232	21 AAA31626	Canine microsatelli
33	25	0.7	252	18 AAT65064	Plant genomic mic
34	25	0.7	252	20 AAX17801	Microsatellite rep
35	25	0.7	341	21 AAF13992	Human foetal liver
36	25	0.7	462	22 ABA55668	Probe #3829 for ge
37	25	0.7	462	22 ABA35363	Human bone marrow
38	25	0.7	462	22 AAK29365	Probe #3881 for ge
39	25	0.7	462	22 AAI13948	Arabidopsis thalia
40	25	0.7	972	21 AAC49195	Arabidopsis thalia
41	25	0.7	1220	21 AAC36229	Arabidopsis thalia
C 42	25	0.7	1676	21 AAC36166	Arabidopsis thalia
C 43	25	0.7	2265	21 AAC43317	Arabidopsis thalia
44	25	0.7	2286	14 AAQ45913	Rice starch synthase
45	25	0.7	2340	23 AAS86699	DNA encoding novel

ALIGNMENTS

RESULT 1
AAC74569
ID AAC74569 standard; cDNA; 333 BP.
XX AAC74569;
AC AAC74569;
DT 08-FEB-2001 (first entry)
XX Human ORFX ORF124 polynucleotide sequence SEQ ID NO:247.

Human: open reading frame; ORFX; detection; cytostatic; hepatotropic; vulnary; antiposoriatic; antiparkinsonian; neurotropic; neuroprotective; anticonvulsant; osteopathic; antirheumatic; immunosuppressive; cardiant; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antinflammatory; antiviral; antibacterial; antifungal; antitumor; antihypertensive; antianemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria; autoimmune disorder; asthma; allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antinflammatory disease; coagulation; thrombosis; contraceptive; ss.

Homo sapiens.

WO200058473-A2.

05-OCT-2000.

31-MAR-2000; 2000WO-US08621.

```

PF 03-APR-1998; 98WO-US06660.
XX
XX 04-APR-1997; 97US-0042939.
PR
XX
XX (EXSE-) EXSEED GENETICS LLC.
PA
XX Guan H, Keeling PL;
PI
XX WPI; 1998-568285/48.
XX P-PSDB; AAY16604.
DR
XX
XX Producing non-glycogen-like polysaccharides in bacteria, fungi or
PT plants - transformed with genes for enzymes involved in starch or
PT glycogen synthesis allows fermentative production of starches with
PT engineered properties
PT
XX Disclosure; Fig 49; 150pp; English.
XX
XX The specification describes a method for the production of
CC non-glycogen-like polysaccharides in a host. The method comprises
CC transforming a host, suitable for fermentation, with genes encoding
CC starch- or glycogen-synthesis enzymes, and fermenting the transformants.
CC The specification also describes hosts transformed with a gene active
CC in glycogen synthesis and at least one non-starch branching gene,
CC involved in production of amylopectin or amylose in its original host.
CC The method is used to produce plant-like starches by fermentation and
CC new starches in plants. These starches are useful for all food and
CC non-food applications of starch. The present sequence is used in
CC the course of the invention.
XX Sequence 1915 BP; 365 A; 610 C; 641 G; 299 T; 0 other;
SQ
Query Match 1.0%; Score 39; DB 19; Length 1915;
Best Local Similarity 100.0%; Pred. No. 3.3e-08;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY 3576 cgctagcacgactacaaggagcgctgggacacagcgtc 3614
Db |||||||
355 cgctagcacgactacaaggagcgctgggacacagcgtc 393
RESULT 3
AAx63355
ID AAx63355 standard; cDNA; 2267 BP.
AC AAx63355;
XX
XX 16-JUL-1999 (first entry)
DT Granule bound starch synthase encoding cDNA.
DE
XX Maize; corn; Zea mays; delta-9 desaturase; GBSS; target; substrate;
KW granule bound starch synthase; hammerhead ribozyme; hairpin ribozyme;
KW modulation; gene expression; transgenic plant; cleavage; canola plant;
KW caffeine synthesis; coffee plant; nicotine production; tobacco;
KW fruit ripening; flower pigmentation; lignin production; ss.
XX
XX Zea mays.
OS
XX WO9710328-A2.
PN
XX 20-MAR-1997.
PD
XX 12-JUL-1996; 96WO-US11699.
PP
XX 13-JUL-1995; 95US-0001135.
PR
XX (DOWC ) DOWELANCO.
XX (RIBO-) RIBOZYME PHARM INC.
FA Edington BE, Folkerts O, Guo L, McSwiggen JA, Merlo DJ;
FA Merlo PAO, Skokut TA, Young SA, Zwick MG;
PI

```



```

FT misc_feature /number= 10
FT 3358..3489 /tag= hh
FT /note= "target duplication site (Ac9)"
FT intron /tag= ii
FT 3395..3489 /number= 10
FT 3490..3681 /tag= jj
FT /number= 11
FT /tag= kk
FT 3570..3572 /number= 12
FT /tag= ll
FT /note= "target duplication site (Spm 18)"
FT intron /tag= ll
FT /number= 11
FT 3793..3879 /tag= mm
FT /number= 12
FT 3880..3976 /tag= nn
FT /number= 12
FT 3977..4105 /tag= oo
FT /number= 13
FT 4106..4226 /tag= pp
FT /number= 13
FT 4227..4595 /tag= qq
FT /number= 14
FT 4570..4575 /tag= rr
FT 4593..4598 /tag= ss
FT /tag= tt
FT polyA_signal
FT polyA_signal
FT polyA_signal
FT WO9814601-A1.
XX
XX
XX 09-APR-1998.
XX 30-SEP-1997; 97WO-US17555.
XX 30-SEP-1996; 96US-0026855.
XX (EXSE-) EXSEED GENETICS LLC.
XX Guan H, Keeling P;
XX WPI: 1998-240100/21.
XX P-PSDB; ANW56484.
XX Hybrid polypeptide comprising starch-encapsulating region and
XX protein - useful for, e.g. producing protein(s) resistant to
XX degradation by stomach acids
XX
XX Example 2; Page 29-31; 156pp; English.
XX
XX The sequence is that of the waxy gene which codes for glucosyl
XX transferase. It can be used in the production of a hybrid
XX polypeptide comprising a starch-encapsulating region (SER) fused
XX to a payload protein. The hybrid polypeptide can be used to make
XX modified starches comprising the payload protein, selected from,
XX e.g. hormones, growth factors, antibodies, enzymes, dyes,
XX immunoglobulins, etc. The modified starch can also be used
XX to provide grain feeds enriched in amino acids. By encapsulating
XX the payload protein in starch, it is more resistant to
XX degradation by stomach acids.
XX
XX Sequence 4800 BP; 935 A; 1414 C; 1446 G; 1005 T; 0 other;

Query Match 1.0%; Score 39; DB 19; Length 4800;
Best Local Similarity 100.0%; Pred. No. 3.3e-08;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3576 cgctagaccagtagacagcagcctgggacaccagcgc 3614
Db 1718 cgctagaccagtagacagcagcctgggacaccagcgc 1756

RESULT 5
ABA68266/C
ID ABA68266 standard; DNA; 276 BP.
XX
XX ABA68266;
XX
XX 01-FEB-2002 (first entry)
XX Human foetal liver single exon nucleic acid probe #16571.
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX Homo sapiens.
XX WO200157277-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00669.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI: 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver -
XX
XX Claim 4; SEQ ID NO 16571; 639pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX foetal liver. The present sequence is a single exon nucleic acid
XX probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 276 BP; 23 A; 81 C; 20 G; 152 T; 0 other;

Query Match 1.0%; Score 37; DB 22; Length 276;
Best Local Similarity 100.0%; Pred. No. 3e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2379 gaagaagaagaagaagaagaagaagaagaagaagaag 2415
Db 243 GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 207

RESULT 6
ABA35275/C
ID ABA35275 standard; DNA; 276 BP.

```


CC present in interval 6D and/or 6E of the distal portion of Yq, mutations
CC in which are associated with reduced sperm count. The DAZH gene
CC (AAZ92580) is located on chromosome 3; however, the entire DAZ gene
CC family, including DAZH is expressed in germ cells. DAZ and DAZH
CC nucleotide sequences may be used as a source of primers and probes for
CC the diagnosis of cases of reduced sperm count associated with alteration
CC or deletion of the DAZ gene. They are also used as human chromosome Y
CC markers. Functional DAZ genes can be used in gene therapy for treating
CC reduced sperm counts. Sequence AAZ92499 represents human DAZ cDNA, and
CC sequences AAZ92583-292584 are genomic DAZ sequences. Sequences
CC AAZ92491-292492 are partial DNA sequences of DAZ gene family members.
XX
SQ Sequence 43795 BP; 12175 A; 8166 C; 8183 G; 15271 T; 0 other;

Query Match 0.8%; Score 31; DB 21; Length 43795;
Best Local Similarity 100.0%; Pred. No. 0.00023;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2385 gaagaagaagaagcagagaagaagaag 2415
|||||
Db 38989 gaagaagaagaagcagagaagaagaag 39019

RESULT 15

AAAC49802
ID AAC49802 standard; DNA; 509 BP.

AC AAC49802;

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 62490.

KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

PN EPI033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.

PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151086.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 22-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.

PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 0.8%; Score 29; DB 21; Length 509;
 Best Local Similarity 100.0%; Pred. No. 0.0021;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2383 aagaagaagaagaagaagaagaagaagaaga 2411
 |||||
 Db 115 aagaagaagaagaagaagaagaagaagaaga 143

Search completed: July 31, 2002, 16:20:22
 Job time: 8259 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 12:26:27 ; Search time 165.21 Seconds
(without alignments)
26.762 Million cell updates/sec

Title: US-09-899-718A-8
Perfect score: 18
Sequence: 1 atactctggcattgttaa 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, NA: *
1: /cgn2.6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2.6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2.6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2.6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2.6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2.6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	14.4	80.0	890	3	US-08-741-411-4
2	14.4	80.0	1070	1	US-08-602-713-11
3	14.4	80.0	1070	4	US-08-989-493-11
4	14.4	80.0	1410	3	US-08-147-592A-1
5	14.4	80.0	1410	4	US-08-292-694A-1
c 6	14.4	80.0	3533	2	US-08-476-062A-40
c 7	14.4	80.0	3533	5	PCT-US96-01314-40
c 8	14.4	80.0	3533	6	5424399-1
9	13.8	76.7	248	4	US-08-991-789A-204
10	13.8	76.7	248	4	US-09-062-451-204
11	13.8	76.7	908	4	US-09-457-046B-15
12	13.8	76.7	3147	4	US-09-153-804-12
13	13.8	76.7	3180	1	US-08-480-662-1
14	13.8	76.7	3180	3	US-08-918-190-1
15	13.8	76.7	3180	4	US-09-234-232-1
16	13.8	76.7	3180	5	PCT-US96-09927-1
c 17	13.8	76.7	3930	2	US-09-018-787-2
c 18	13.8	76.7	12730	4	US-09-004-838-91
c 19	13.8	76.7	15062	4	US-09-004-838-89
20	13.4	74.4	17	1	US-08-373-124A-1713
21	13.4	74.4	17	1	US-08-435-628-1713
c 22	13.4	74.4	825	1	US-08-307-499-48
c 23	13.4	74.4	825	4	US-09-299-268-48
c 24	13.4	74.4	1234	1	US-08-798-000-1
c 25	13.4	74.4	1328	4	US-09-475-316A-71
c 26	13.4	74.4	6028	4	US-09-011-745-5
27	13.4	74.4	8202	1	US-08-258-420-13

28	13.4	74.4	14176	1	US-08-307-499-1	Sequence 1, Appli
c 29	13.4	74.4	14176	1	US-08-307-499-14	Sequence 14, Appl
30	13.4	74.4	14176	4	US-09-299-268-1	Sequence 1, Appli
c 31	13.4	74.4	14176	4	US-09-299-268-14	Sequence 14, Appl
c 32	13.2	73.3	24	2	US-08-680-327-6	Sequence 6, Appli
33	13.2	73.3	132	1	US-08-248-474-5	Sequence 5, Appli
34	13.2	73.3	132	3	US-08-756-849-5	Sequence 5, Appli
c 35	13.2	73.3	455	6	5278286-1	Patent No. 5278286
c 36	13.2	73.3	601	2	US-08-184-009-168	Sequence 168, App
c 37	13.2	73.3	601	2	US-08-458-356-168	Sequence 168, App
c 38	13.2	73.3	601	4	US-08-460-736-168	Sequence 168, App
39	13.2	73.3	736	3	US-08-771-110-1	Sequence 1, Appli
c 40	13.2	73.3	1095	2	US-08-625-377-3	Sequence 3, Appli
c 41	13.2	73.3	1095	4	US-09-204-841-3	Sequence 1, Appli
c 42	13.2	73.3	1251	4	US-09-242-095-1	Sequence 3, Appli
43	13.2	73.3	1452	1	US-08-149-093A-3	Sequence 3, Appli
44	13.2	73.3	1452	1	US-08-911-245-3	Sequence 3, Appli
45	13.2	73.3	1452	1	US-08-553-058C-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-741-411-4/c
; Sequence 4, Application US/08741411
; Patent No. 6124116
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: NOVEL RAB PROTEINS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/741.411
; FILING DATE: Herewith
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0139 US
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 890 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; LIBRARY: Consensus
; CLONE: US-08-741-411-4

Query Match 80.0%; Score 14.4; DB 3; Length 890;

Best Local Similarity 93.8%; Pred. No. 47;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atactctggtcatgtt 16
||||| |||||||
Db 282 ATACTCGGTCATGTT 267

RESULT 2

US-08-602-713-11
; Sequence 11, Application US/08602713
; Patent No. 5798205
; GENERAL INFORMATION:
; APPLICANT: Hauser, Hans-Peter; Knapp Stefan; Brust, Stefan;
; APPLICANT: G rtler, Lutz G.; Eberle, Josef; Kaptue, Lazare;
; APPLICANT: Zekeng, L opold Achengu
; TITLE OF INVENTION: Retrovirus From The HIV Group And Its Use
; TITLE OF INVENTION: (MVP-2901/94)
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felie & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/602,713
; FILING DATE: 16-FEBRUARY-1996
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 195 05 262
; FILING DATE: 16-FEBRUARY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5798205man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LEDE 203
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884

; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1070 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
US-08-602-713-11

Query Match 80.0%; Score 14.4; DB 1; Length 1070;
Best Local Similarity 93.8%; Pred. No. 48;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atactctggtcatgtt 16
||||| |||||||
Db 403 ATACTCTGTCATGTT 418

RESULT 3

US-08-989-493-11
; Sequence 11, Application US/08989493
; Patent No. 6162631
; GENERAL INFORMATION:
; APPLICANT: Hauser, Hans-Peter; Knapp Stefan; Brust, Stefan;
; APPLICANT: G rtler, Lutz G.; Eberle, Josef; Kaptue, Lazare;
; APPLICANT: Zekeng, L opold Achengu
; TITLE OF INVENTION: Retrovirus From The HIV Group And Its Use

; TITLE OF INVENTION: (MVP-2901/94)
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felie & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/989,493
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/602,713
; FILING DATE: 16-FEBRUARY-1996
; APPLICATION NUMBER: DE 195 05 262
; FILING DATE: 16-FEBRUARY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 6162631man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LEDE 203
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884

; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1070 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
US-08-989-493-11

Query Match 80.0%; Score 14.4; DB 4; Length 1070;
Best Local Similarity 93.8%; Pred. No. 48;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atactctggtcatgtt 16
||||| |||||||
Db 403 ATACTCTGTCATGTT 418

RESULT 4

US-08-147-592A-1
; Sequence 1, Application US/08147592A
; Patent No. 6096513
; GENERAL INFORMATION:
; APPLICANT: Bell, Graeme I
; APPLICANT: Reisine, Terry
; APPLICANT: Yasuda, Kazuki
; TITLE OF INVENTION: Opioid Receptor Genes,
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America
; ZIP: 72210

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```

1  APPLICATION NUMBER:  US/08/147,592A
2  FILING DATE:  05-NOV-1993
3  CLASSIFICATION:  435
4  PRIOR APPLICATION DATA:
5  APPLICATION NUMBER:  08/100,694
6  FILING DATE:  30-JUL-1993
7  CLASSIFICATION:  435
8  APPLICATION NUMBER:  08/066,296
9  FILING DATE:  20-MAY-1993
10 CLASSIFICATION:  435
11 ATTORNEY/AGENT INFORMATION:
12 NAME:  Wilson, Mark B.
13 REGISTRATION NUMBER:  37,259
14 REFERENCE/DOCKET NUMBER:  ARCD:105
15 TELECOMMUNICATION INFORMATION:
16 TELEPHONE:  (512) 418-3000
17 TELEFAX:  (512) 474-7577
18 TELEX:  N/A
19 INFORMATION FOR SEQ ID NO:  1:
20 SEQUENCE CHARACTERISTICS:
21 LENGTH:  1410 base pairs
22 TYPE:  nucleic acid
23 STRANDEDNESS:  single
24 TOPOLOGY:  linear
25 MOLECULE TYPE:  DNA
26 FEATURE:
27 NAME/KEY:  CDS
28 LOCATION:  186..1325
29 US-08-147-592A-1

```

```

Query Match      80.0%; Score 14.4; DB 3; Length 1410;
Best Local Similarity 93.8%; Pred. NO. 50;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

Qy 1 atactctgggtcatgtt 16
 ||| ||||| ||||| |||
Db 415 ATTCTCTGGTCATGTT 430

```

RESULT      5
US-08-292-694A-1
: Sequence 1, Application US/08292694A
: Patent No. 6319686
: GENERAL INFORMATION:
: APPLICANT: BELI, GRAEME
: APPLICANT: REISINE, TERRY
: APPLICANT: YASUDA, KAZUKI
: TITLE OF INVENTION: OPIOID RECEPTORS: COMPOSITIONS AND METHODS
: NUMBER OF SEQUENCES: 46
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Arnold, White & Durkee
: STREET: P. O. Box 4433
: CITY: Houston
: STATE: Texas
: COUNTRY: USA
: ZIP: 77210
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/292,694A
: FILING DATE: August 19, 1994
: CLASSIFICATION: 435
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: 08/066,296
: FILING DATE: 20 May 1993
: CLASSIFICATION: 435
: APPLICATION NUMBER: 08/100,694
: FILING DATE: 30 July, 1993
: CLASSIFICATION: 435

```

```

1 APPLICATION NUMBER: 08/147,592
2 FILING DATE: 5 NO. 6319686member 1993
3
4 CLASSIFICATION: 435
5 APPLICATION NUMBER: PCT/US94/05747
6 FILING DATE: 20 May 1994
7
8 CLASSIFICATION: 435
9 ATTORNEY/AGENT INFORMATION:
10 NAME: MARK B. WILSON
11 REGISTRATION NUMBER: 37,259
12 REFERENCE/DOCKET NUMBER: ARCD:140/WIM
13 TELECOMMUNICATION INFORMATION:
14 TELEPHONE: (512) 418-3000
15 TELEFAX: (713) 789-2679
16 TELEX: 79-0924
17
18 INFORMATION FOR SEQ ID NO: 1:
19 SEQUENCE CHARACTERISTICS:
20 LENGTH: 1410 base pairs
21 TYPE: nucleic acid
22 STRANDEDNESS: single
23 TOPOLOGY: linear
24 MOLECULE TYPE: genomic DNA
25 FEATURE:
26 NAME/KEY: CDS
27 LOCATION: 186..1325
28 US-08-292-694A-1

```

Query Match	80.0%;	Score 14.4;	DB 4;	Length 1410;
Best Local Similarity	93.8%;	Pred. No. 50;		
Matches	15;	Conservative	0;	Mismatches
			1;	Indels
			0;	Gaps
				0;

Qy 1 atactctgggtcatgtt 16
|||
Db 415 ATTCTCTGGTCATGTT 430

RESULT 6
US-08-476-062A-40/c
; Sequence 40, Application US/08476062A
; Patent No. 5877275
; GENERAL INFORMATION:
; APPLICANT: Arnaout, M. Amin
; TITLE OF INVENTION: CONTROLLING CELLULAR IMMUNE/INFLAMMATORY
; ; TITLE OF INVENTION: RESPONSES WITH BETA2 INTEGRINS
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/476,062A
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/216,081
; FILING DATE: 21-MAR-1994
; APPLICATION NUMBER: 07/637,830
; FILING DATE: 04-JAN-1991
; APPLICATION NUMBER: 07/539,842
; FILING DATE: 18-JUN-1990
; APPLICATION NUMBER: 07/212,573
; FILING DATE: 28-JUN-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Freeman, John W.
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 00796/068003

Query Match 80.0%; Score 14.4; DB 5; Length 3533;
Best Local Similarity 93.8%; Pred. NO. 56;
Matches 15; Conservative 0; Mismatches 1; Indels 0.

TELEPHONE: (206) 622-49
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 204:
SEQUENCE CHARACTERISTICS:

```
Query Match      80.0%; Score 14.4; DB 5; Length 3533;
Best Local Similarity 93.8%; pred. No. 56;
Matches 15: Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```


us-09-899-718a-8.rni

Thu Aug 1 08:30:07 2002

```
; FILE REFERENCE: 53679
; CURRENT APPLICATION NUMBER: US/09/457,046B
; CURRENT FILING DATE: 1999-12-07
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 908
; TYPE: DNA
; ORGANISM: Taxus cuspidata
; US-09-457-046B-15

Query Match          76.7%; Score 13.8; DB 4; Length 908;
Best Local Similarity 88.2%; Pred. No. 99;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atactctggtcatgtta 17
   || ||||| ||||| |||||
Db 71 ATGCTCTGATCATGTGA 87

RESULT 10
US-09-062-451-204
; Sequence 204, Application US/09062451
; Patent No. 6344550
; GENERAL INFORMATION:
; APPLICANT: Fridakis, Tony N.
; APPLICANT: Smith, John M.
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF BREAST CANCER
; NUMBER OF SEQUENCES: 297
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/062,451
; FILING DATE: 04-APR-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.419C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 204:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 248 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-062-451-204

Query Match          76.7%; Score 13.8; DB 4; Length 248;
Best Local Similarity 88.2%; Pred. No. 83;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atactctggtcatgtta 17
   || ||||| ||||| |||||
Db 71 ATGCTCTGATCATGTGA 87

RESULT 11
US-09-457-046B-15
; Sequence 15, Application US/09457046B
; Patent No. 6287835
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney et al.
; TITLE OF INVENTION: Transacylases of the Paclitaxel Biosynthetic Pathway

; FILE REFERENCE: 53679
; CURRENT APPLICATION NUMBER: US/09/457,046B
; CURRENT FILING DATE: 1999-12-07
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 908
; TYPE: DNA
; ORGANISM: Taxus cuspidata
; US-09-457-046B-15

Query Match          76.7%; Score 13.8; DB 4; Length 908;
Best Local Similarity 88.2%; Pred. No. 99;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atactctggtcatgtta 17
   || ||||| ||||| |||||
Db 663 atactatggttaagtta 679

RESULT 12
US-09-153-804-12
; Sequence 12, Application US/09153804
; Patent No. 6207380
; GENERAL INFORMATION:
; APPLICANT: Patricia Billing-Medel
; APPLICANT: Maurice Cohen
; APPLICANT: Paula N. Friedman
; APPLICANT: Julian Gordon
; APPLICANT: Steven C. Hodges
; APPLICANT: Michael R. Klass
; APPLICANT: Jon D. Kratochvil
; APPLICANT: Eric Russell
; APPLICANT: Stephen D. Stroupe
; TITLE OF INVENTION: Reagents and Methods Useful for Detecting Diseases of the Urin
; TITLE OF INVENTION: Tract
; FILE REFERENCE: 6180.US.01
; CURRENT APPLICATION NUMBER: US/09/153,804
; CURRENT FILING DATE: 1998-09-15
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 12
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-153-804-12

Query Match          76.7%; Score 13.8; DB 4; Length 3147;
Best Local Similarity 88.2%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tactctggtcatgttaa 18
   | ||||| ||||| |||||
Db 1336 ttctctggttaagttaa 1352

RESULT 13
US-08-480-662-1
; Sequence 1, Application US/08480662
; Patent No. 5759782
; GENERAL INFORMATION:
; APPLICANT: Pastan, Ira
; APPLICANT: Brinkmann, Ulrich
; TITLE OF INVENTION: CELLULAR APOPTOSIS SUSCEPTIBILITY PROTEIN (CSP) AND AN
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson and Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: USA
```

```
;
;
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,662
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Israel, Ned A
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH12.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-235-8550
; TELEFAX: 619-235-0176
; TELEX:
;
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3180 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; US-08-480-662-1

Query Match 76.7%; Score 13.8; DB 1; Length 3180;
Best Local Similarity 88.2%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 tactctggatcatgtaa 18
Db 1336 TTCTCTGGTATGTTAA 1352

RESULT 14
US-08-918-190-1
; Sequence 1, Application US/08918190
; Patent No. 6072031
; GENERAL INFORMATION:
; APPLICANT: Pastan, Ira
; TITLE OF INVENTION: CELLULAR APOPTOSIS SUSCEPTIBILITY
; TITLE OF INVENTION: PROTEIN (CSP) AND ANTISENSE CSP
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson and Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: USA
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/918,190
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/480,662
; FILING DATE: 07-JUN-1995
```

```
;
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Israel, Ned A
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH12.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-235-8550
; TELEFAX: 619-235-0176
; TELEX:
;
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3180 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; US-08-918-190-1

Query Match 76.7%; Score 13.8; DB 3; Length 3180;
Best Local Similarity 88.2%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 tactctggatcatgtaa 18
Db 1336 TTCTCTGGTATGTTAA 1352

RESULT 15
US-09-234-232-1
; Sequence 1, Application US/09234232
; Patent No. 6232086
; GENERAL INFORMATION:
; APPLICANT: Pastan, Ira
; TITLE OF INVENTION: CELLULAR APOPTOSIS SUSCEPTIBILITY
; TITLE OF INVENTION: PROTEIN (CSP) AND ANTISENSE CSP
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson and Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: USA
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/234,232
; FILING DATE: 20-Jan-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/918,190
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Israel, Ned A
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH12.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-235-8550
; TELEFAX: 619-235-0176
; TELEX: <Unknown>
;
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3180 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
```

us-09-899-718a-8.rni

Thu Aug 1 08:30:07 2002

```

;
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: <Unknown>
; ORIGINAL SOURCE:
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-234-232-1

Query Match      76.7%; Score 13.8; DB 4; Length 3180;
Best Local Similarity 88.2%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      2 tactctggtcatgttaa 18
        | | | | | | | | | |
Db 1336 TTCCTCTGTTATGTAA 1352

Search completed: July 31, 2002, 12:26:28
Job time: 12011 sec

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 12:12:12 ; Search time 5855.71 Seconds
(without alignments)
41.489 Million cell updates/sec

Title: US-09-899-718A-8
Perfect score: 18
Sequence: 1 atactctgtcatgttaa 18

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_estin:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pin:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	17	94.4	1441	10	BG531376	BG531376 602559625
C 2	16.4	91.1	344	10	BG516695	BG516695 NXSL 002
C 3	16.4	91.1	513	10	BG318395	BG318395 NXPV 013
C 4	16.4	91.1	558	12	A0621441	A0621441 HS_3019_B
C 5	16.4	91.1	564	12	A2144687	A2144687 SP_0041_B
C 6	16.4	91.1	571	10	BG318023	BG318023 NXPV 008
C 7	16.4	91.1	724	12	A2207084	A2207084 SP_0127_A
C 8	16	88.9	404	12	A0150695	A0150695 HS_3203_A
C 9	16	88.9	635	9	AW812504	AW812504 CM4-ST018
C 10	15.6	86.7	1026	12	CNS0275Q	AL1212759 Tetraodon
C 11	15.4	85.6	213	10	BE1287040	BE1287040 UI-R-Cf05
C 12	15.4	85.6	225	9	BE142340	BE142340 CM3-HT014
C 13	15.4	85.6	297	9	BB024447	BB024447 BB024447
C 14	15.4	85.6	303	9	BB289468	BB289468 BB289468
C 15	15.4	85.6	337	10	BF147142	BF147142 uy07f02.Y
C 16	15.4	85.6	350	10	BE556305	BE556305 sq01a05.Y
C 17	15.4	85.6	393	10	BE650684	BE650684 UI-M-BH3

C 18	15.4	85.6	410	12	AQ442997	AQ442997 HS_5132_B
C 19	15.4	85.6	425	9	AA102952	AA102952 mo09d01.r
C 20	15.4	85.6	443	10	BF256005	BF256005 HVSMEF000
C 21	15.4	85.6	451	10	BI326958	BI326958 AR071D101
C 22	15.4	85.6	457	9	AW738969	AW738969 gb16h09.Y
C 23	15.4	85.6	476	9	AA152987	AA152987 mq48c03.r
C 24	15.4	85.6	481	9	AI493734	AI493734 qz12g07.x
C 25	15.4	85.6	494	9	AW230662	AW230662 uo66d09.Y
C 26	15.4	85.6	495	12	AQ427545	AQ427545 CITBI-E1-
C 27	15.4	85.6	505	10	HI6901	HI6901 ym39p02.r1
C 28	15.4	85.6	512	9	AW395872	AW395872 sh01c11.Y
C 29	15.4	85.6	529	10	C84686	C84686 C84686 Dict
C 30	15.4	85.6	537	12	AQ596806	AQ596806 HS_5206_A
C 31	15.4	85.6	543	10	BI784821	BI784821 saf93q09.
C 32	15.4	85.6	552	12	A2446415	A2446415 IM0242014
C 33	15.4	85.6	558	10	BI784791	BI784791 saf93g11.
C 34	15.4	85.6	564	9	AW232748	AW232748 f129g04.x
C 35	15.4	85.6	571	12	AQ802955	AQ802955 HS_3160_A
C 36	15.4	85.6	617	12	AQ787071	AQ787071 HS_3156_B
C 37	15.4	85.6	629	12	AG017311	AG017311 Homo sapi
C 38	15.4	85.6	634	12	AG058023	AG058023 Pan trogl
C 39	15.4	85.6	635	10	BE369456	BE369456 601220868
C 40	15.4	85.6	635	12	AQ018921	AQ018921 CIT-HSP-2
C 41	15.4	85.6	637	12	AG017310	AG017310 Homo sapi
C 42	15.4	85.6	646	12	AG017312	AG017312 Homo sapi
C 43	15.4	85.6	673	12	BH094331	BH094331 RPCI-24-3
C 44	15.4	85.6	688	10	BG694687	BG694687 NISC-1v06
C 45	15.4	85.6	697	9	AV836265	AV836265 AV836265

ALIGNMENTS

RESULT 1
BG531376/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BG531376 1441 bp mRNA linear EST 03-APR-2001
602559625F1 NIH_MGC_61 Homo sapiens cDNA clone IMAGE:4697383 5',
mRNA sequence.
BG531376
BG531376.1 GI:13522913
EST.
HUMAN.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1441)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC

CDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

Plate: LLCM1526 row: b column: 08
High quality sequence start: 186
High quality sequence stop: 226.
Location/Qualifiers
1. 1441

FEATURES
source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4697383"
/clone.lib="NIH-MGC.61"
/tissue_type="embryonal carcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: testis; Vector: pDNR-LIB (Clontech); Site:1:
SfiI (ggccgcctcgcc); Site:2: SfiI (ggccattatggcc);
Double-stranded cDNA was prepared from cell line RNA.
and 3' adaptors were used in cloning as follows: 5'

adaptor sequence: 5'-CACGCGATATGCC-3' and 3' adaptor sequence: 5'-ATTGACGCGGCGGCGACATG-ctt(30)BN-3' (where B = A, C, G and N = A, C, G, T). Average insert size 1.75 kb (range 0.9-4.0 kb), 15/12 clones contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH-MCC Library."

BASE COUNT	413 a	381 c	367 g	279 t	1 others
ORIGIN					
Query Match			94.4%	Score 17;	DB 10;
Best Local Similarity			100.0%;	Pred. No. 3e+02;	Length 1441;
Matches 17;			Conservative	0;	Mismatches 0;
				Indels	0;
				Gaps	0;

Qy	1	atactctggtcatgcta	17
Db	183	ATACTCTGGTCATGTTA	167
RESULT	2		
BF516695			
LOCUS	344 bp mRNA linear EST 08-DEC-2000		
DEFINITION	NXSI_002_C09_F NXSI (Nsf Xylem side wood Inclined) Pinus taeda cDNA clone NXSI_002_C09_5', mRNA sequence.		

ACCESSION	BF516695
VERSION	BF516695.1
KEYWORDS	GI:11603809
EST.	
SOURCE	loblolly pine.
ORGANISM	Pinus taeda
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus. 1 (bases 1 to 344)
AUTHORS	Sederoff, R.

```

FEATURES
seq primer=13:
source
Location/Qualifiers
1..344
/organism="Pinus taeda"
/strain="Coastal plain loblolly pine from North Carolina"
/db_xref="taxon:3352"
/clone="NXSI_002_C09"
/clone_lib="NXSI"
/tissue_type="xylem"
/cell_type="side"

```

BASE COUNT

80 a	74 c	75 q	103 t	12 others
------	------	------	-------	-----------

```
/dev_stage="Juvenile"  
/lab_host="XLI-Blue"  
/note="vector: Bluescript SK; Site_1: Eco RI; Site_2: XhoI  
; The library is from early (spring) wood, taken from  
three six-year old trees (three different genotypes), in  
the juvenile phase. These trees were induced to form side  
wood by bending to a 45 degree angle and tying them to the  
ground. Differentiating xylem was harvested from the sides  
of the inclined stems, and a mixture of all three  
genotypes was used for the library. oligo-dT primed cDNA  
was directionally cloned into the EcoRI-XhoI Bluescript SK  
vector arms. NOTE: The sequences contain a 'cDNA adapter'  
between the EcoRI site and the start of the EST. The  
adapter sequence is 'AATTGGCAGCAG'."
```

Qy 1 atactctggtcatgttaa 18
 |||||
 Db 237 ATACACTGGTCATGTTAA 254

RESULT	3
BG318395	
LOCUS	
DEFINITION	EG318395 513 bp mRNA linear EST 26-FEB-2001 NXPV_013_B06_F NXPV (Nsf Xylem Planings wood Vertical) Pinus taeda cDNA clone NXPV_013_B06_5', mRNA sequence.

ACCESSION	BG318395	
VERSION	BG318395.1	GI:13127
KEYWORDS	EST.	
SOURCE		

SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	COMMENT
Isoliolly pine.	Pinus taeda	Eukaryota; Viridipla		Molecular Basis of W	
		Spermatophyta; Conif		Unpublished (2000)	
		1 (bases 1 to 513)	Sederoff, R.		Contact: Johnson, Ar

North Carolina State
Tel: 919 515 7800
Fax: 919 515 7801

```

FEATURES
source
Location/Qu
1. .513
/organism="
/organism="Co
Seq primer: T3.
Email: ajohnson@unit

```

REFERENCE 1 (bases 1 to 558)
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589

COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones may be purchased from Research Genetics (info@resgen.com).
 BAC end Web Server: http://www.htsc.washington.edu

Plate: 3019 row: L column: 10

Seq primer: 17

Class: BAC ends

High quality sequence stop: 558.

Location/Qualifiers

1..558

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="Plate-3019 Col-10 Row=L"

/clone_lib="CIR Approved Human Genomic Sperm Library D"

/sex="male"

/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 145 a 142 c 118 g 150 t 3 others

ORIGIN

Query Match 91.1%; Score 16.4; DB 12; Length 558;

Best Local Similarity 94.4%; Pred. No. 4.6e+02;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atactctggtcattgttaa 18

||||| |||||||

Db 437 ATACTCTGCTCATGTAA 420

RESULT 5

AZ144687

LOCUS

DEFINITION SP_0041_B2_F12.SP6E Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library Strongylocentrotus purpuratus genomic clone Plate=41 Col=24 Row=L, DNA sequence.

ACCESSION AZ144687

VERSION AZ144687.1 GI:8296590

KEYWORDS GSS.

SOURCE Strongylocentrotus purpuratus.

ORGANISM Strongylocentrotus purpuratus

Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;

Echinoidea; Euechinoidea; Echinacea; Echinoida;

Strongylocentrotidae; Strongylocentrotus.

1 (bases 1 to 564)

Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R., Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and Hood,L.

A sea urchin genome project: Sequence scan, virtual map, and additional resources

Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)

20402566

COMMENT Contact: Cameron, RA, Davidson, EH, Hood, L

Division of Biology 156-29

California Institute of Technology

Pasadena California 91125, USA

Tel: (626) 395-8421

Fax: (626) 793-3047

Email: acameron@caltech.edu

Plate: 41 row: L column: 24

Seq primer: SP6

Class: BAC ends

High quality sequence stop: 564.

Location/Qualifiers

1..564

/organism="Strongylocentrotus purpuratus"

/db_xref="taxon:7668"

/clone="Plate=41 Col=24 Row=L"

/clone_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library"

/note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli DH10B"

BASE COUNT 188 a 97 c 127 g 146 t 6 others

ORIGIN

Query Match 91.1%; Score 16.4; DB 12; Length 564;

Best Local Similarity 94.4%; Pred. No. 4.6e+02;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atactctggtcattgttaa 18

||||| |||||||

Db 253 ATACTTGGTCATGTAA 270

RESULT 6

BG318023

LOCUS

DEFINITION NXPV_008_F06_F NXPV (Nsf Xylem Planings wood vertical) Pinus taeda

CDNA clone NXPV_008_F06 5', mRNA sequence.

ACCESSION BG318023

VERSION BG318023.1 GI:13127453

KEYWORDS EST.

SOURCE lobliolly pine.

ORGANISM Pinus taeda

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.

1 (bases 1 to 571)

Sederoff,R.

Molecular Basis of Wood Formation in the Pine Megagenome

Unpublished (2000)

Contact: Johnson, Arthur

North Carolina State University

Tel: 919 515 7800

Fax: 919 515 7801

Email: ajohnson@unity.ncsu.edu

Seq primer: T3.

Location/Qualifiers

1..571

/organism="Pinus taeda"

/strain="Coastal plain lobliolly pine from North Carolina"

/db_xref="taxon:3352"

/clone="NXPV_008_F06"

/clone_lib="NXPV (Nsf Xylem Planings wood vertical)"

/tissue_type="Xylem"

/cell_type="Planings (secondary)"

/dev_stage="Transitional"

/lab_host="XLI-Blue"

/note="Vector: Bluescript SK; Site_1: Eco RI; Site_2: XhoI

; The library is from early (Spring) secondary wood, taken from a ten year old tree in the transitional phase. The tree is a kind gift of the Westvaco Corporation. Secondary

xylem was harvested from the tree by peeling back the bark and primary xylem and then removing the underlying tissue with a block plane. NOTE: The sequences contain a 'cDNA

adapter' between the EcoRI site and the start of the EST.

The adapter sequence is 'AATTGGCAGAG'.

BASE COUNT 141 a 111 c 129 g 169 t 21 others

ORIGIN

Query Match 91.1%; Score 16.4; DB 10; Length 571;

```

Best Local Similarity 94.4%; Pred. No. 4.7e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 atactctggtcatgttaa 18
    ||||| ||||| ||||| |||||
Db 324 ATACACTGGTCATGTAA 341

RESULT 7
AZ207084/c
LOCUS
DEFINITION
SP_0127_A2_G12_T7A Strongylocentrotus purpuratus, purple sea urchin
'sperm' genomic BAC library Strongylocentrotus purpuratus genomic
clone Plate-127 Col-24 Row=M, DNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Strongylocentrotus purpuratus.
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinoidea; Euechinoidea; Echinacea; Echinoidea;
Strongylocentrotidae; Strongylocentrotus.
REFERENCE
AUTHORS
1 (bases 1 to 724)
Swartzell, S., Wallace, J.C., Roustka, A.J., Livingston, B.T., Wray
, G.A., Ettensohn, C.A., Lehrach, H., Britten, R.J., Davidson, E.H. and
Hood, L.
A sea urchin genome project: Sequence scan, virtual map, and
additional resources
Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)
20402566
Contact: Cameron, RA, Davidson, EH, Hood, L
Division of Biology 156-29
California Institute of Technology
Pasadena California 91125, USA
Tel: (626) 395-8421
Fax: (626) 793-3047
Email: acameron@caltech.edu
Plate: 127 Row: M Column: 24
Seq primer: T7
Class: BAC ends
High quality sequence stop: 724.
Location/Qualifiers
1. .724
/organism="Strongylocentrotus purpuratus"
/db_xref="taxon:7668"
/clone_lib="Strongylocentrotus purpuratus, purple sea
urchin, sperm genomic BAC library"
/note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli
DH10B"
BASE COUNT
198 a 149 c 103 g 274 t
ORIGIN
source

Query Match 91.1%; Score 16.4; DB 12; Length 724;
Best Local Similarity 94.4%; Pred. No. 5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 atactctggtcatgttaa 18
    ||||| ||||| ||||| |||||
Db 575 ATACTTTGGTCATGTAA 558

RESULT 8
AQ150695
LOCUS
DEFINITION
HS_3203_A2_C12_MR CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate-3203 Col-24 Row=E, DNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
human.

```

```

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS
1 (bases 1 to 404)
Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and
Hood, L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
JOURNAL
MEDLINE
COMMENT
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3203 Row: E Column: 24
Class: BAC ends
High quality sequence stop: 404.
Location/Qualifiers
1. .404
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3203 Col=24 Row=E"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT
109 a 87 c 98 g 110 t
ORIGIN

Query Match 88.9%; Score 16; DB 12; Length 404;
Best Local Similarity 100.0%; Pred. No. 6.8e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atactctggtcatgtt 16
    ||||| ||||| ||||| |||||
Db 285 ATACTCTGGTCATGTT 300

RESULT 9
AW812504/c
LOCUS
DEFINITION
CM4-ST0181-231199-049-b12 ST0181 Homo sapiens cDNA, mRNA sequence.
ACCESSION
VERSION
AW812504.1 GI:7905498
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS
1 (bases 1 to 635)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
JOURNAL
MEDLINE
COMMENT
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

```


Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?ti=&t2=CM4-ST0181-231
 199-049-b12&t3=1999-11-23&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 18
 High quality sequence stop: 586.

FEATURES

Location/Qualifiers
 1..635
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="ST0181"
 /dev_stage="Adult"

/note="Organ: stomach; Vector: puc18; Site_1: SmaI;
 Site_2: SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the pUC 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."
 203 a 101 c 135 g 196 t

BASE COUNT

ORIGIN

Query Match 88.9%; Score 16; DB 9; Length 635;
 Best Local Similarity 100.0%; Pred. No. 7.6e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atactctggtcatgtt 16

Db 94 ATACTCTGGTCATGTT 79

RESULT 10

CNS02T5Q/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

CNS02T5Q 1026 bp DNA linear GSS 15-MAY-2000
 Tetraodon nigroviridis genome survey sequence T7 end of clone
 163022 of library G from Tetraodon nigroviridis, genomic survey
 sequence.

AL212759
 AL212759.1 GI:7871578
 GSS: genome survey sequence.
 Tetraodon nigroviridis.

Tetraodon nigroviridis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes;
 Tetraodontidae; Tetraodon.

1 (bases 1 to 1026)
 Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizes, C., Fisher, C.,
 Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
 Weissenbach, J.

Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetraodon nigroviridis
 Unpublished

2 (bases 1 to 1026)
 Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
 Bernot, A., Fizes, C., Wincker, P., Brottier, P., Quetier, F.,
 Saurin, W. and Weissenbach, J.

Human gene number estimate provided by genome wide analysis using
 Tetraodon nigroviridis DNA sequence
 Unpublished

3 (bases 1 to 1026)
 Genoscope.
 Direct Submission
 Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
 This sequence is a single read and was generated as part of a large
 scale clone-end sequencing project of the Tetraodon nigroviridis
 genome. For more information, please take a look at
 http://www.genoscope.cns.fr/Tetraodon.

Location/Qualifiers
 1..1026

/organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone="163022"
 /clone_lib="G"
 /note="Genoscope sequence ID : COAG163BH11L1-end : T7"
 325 a 201 c 215 g 274 t 11 others

BASE COUNT

ORIGIN

Query Match 86.7%; Score 15.6; DB 12; Length 1026;
 Best Local Similarity 93.8%; Pred. No. 1.4e+03;
 Matches 15; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 actctggtcatgttaa 18

Db 497 ACTCTGGTCATKTTAA 482

RESULT 11

BI287040/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

BI287040 213 bp mRNA linear EST 19-JUL-2001
 UI-R-CT0s-cax-d-05-0-UI.s1 UI-R-CT0s Rattus norvegicus cDNA clone
 UI-R-CT0s-cax-d-05-0-UI.3', mRNA sequence.

BI287040
 BI287040.1 GI:14942270
 EST.
 Norway rat.
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

1 (bases 1 to 213)
 Bonaldo, M.F., Lennon, G. and Soares, M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery

Genome Res. 6 (9), 791-806 (1996)
 97044477
 Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. The sequence tag present in the cDNA between the NotI site
 and the oligo-dT track served to identify it as a clone from the
 non-normalized rat brain pool library cDNA Library Preparation:
 M.B. Soares Lab Clone distribution: clones will be available
 through Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLYA-Yes.

Location/Qualifiers
 1..213
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-CT0s-cax-d-05-0-UI"
 /clone_lib="UI-R-CT0s"
 /dev_stage="ADULT"

/lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CT0s
 library is a non-normalized library constructed from the
 following rat brain tissues: embryonic day 17, embryonic
 day 19, embryonic day 21, adult day 1, adult day 12, adult
 day 75, adult day 200. For a detailed description of the
 library from which this clone was derived, please visit
 our web site at ratest.eng.uiowa.edu. The subtraction has
 been previously described in (Bonaldo, Lennon and Soares,
 Genome Research 6:791-806, 1996)

FEATURES

source

Query Match 85.6%; Score 15.4; DB 10; Length 213;
 Best Local Similarity 94.1%; Pred. No. 1.1e+03;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 TAG_LIB=UI-R-CT0s
 TAG_TISSUE=rat brain pool
 TAG_SEQ=ACATC
 BASE COUNT 67 a 35 c 49 g 62 t
 ORIGIN

2 tactctgtgcatgttaa 18
 |||||
 49 TACTCTGGTCATCTTAA 33

RESULT 12

BE142340
 LOCUS CM3-HT0143-220999-015-b05 HT0143 Homo sapiens mRNA linear EST 21-JUN-2000
 DEFINITION
 ACCESSION BE142340
 VERSION BE142340.1 GI:8605061
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 225)
 Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Balda,G.S., Simpson,D.H.,
 Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,
 M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
 Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE 20202663

COMMENT Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prudente Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2=CM3-HT0143-220

999-015-b05&t3=1999-09-22&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 102

High quality sequence stop: 225.

Location/Qualifiers

1. 225

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="HT0143"

/dev_stage="Adult"

/note="Organ: head_neck; Vector: puc18; Site: 1: Smal;

Site: 2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (U.S. Letters Patent application

No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the puc 18 vector. Reverse transcription of

tissue mRNA and cDNA amplification were performed under

low stringency conditions."

52 a 54 c 55 g 64 t

Query Match 85.6%; Score 15.4; DB 9; Length 225;

Best Local Similarity 94.1%; Pred. No. 1.2e+03;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atactctgtgcatgttaa 17
 |||||
 Db 76 ATACTCTGGTCATCTTAA 92

RESULT 13

BB024447

LOCUS BB024447

DEFINITION BB024447

ACCESSION BB024447

VERSION BB024447

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 297)

Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci

P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,

Hirozane,T., Horii,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,

Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,

Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.

, Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.

, Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata,Y.

, Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomimaga,N., Toya

, Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.

, Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino

M., Muramatsu,M. and Hayashizaki,Y.

RIKEN Mouse ESTs (Konno, H., et al.)

Unpublished (2000)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/

Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki

N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Thermostabilization and thermoactivation of thermolabile enzymes by

trichalose and its application for the synthesis of full length

cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)

Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,

Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki

,Y. and Hayashizaki,Y.

Automated filtration-based high-throughput plasmid preparation

system. Genome Res. 9 (5), 463-470 (1999)

Carninci,P. and Hayashizaki,Y.

High-efficiency full-length cDNA cloning. Methods Enzymol. 303,

19-44 (1999)

Please visit our web site (http://genome.rtc.riken.go.jp) for

further details.

Location/Qualifiers

1. 297

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone_lib="RIKEN full-length enriched, adult male

pituitary gland"

/sex="male"

/tissue_type="pituitary gland"

/dev_stage="adult"

/lab_host="DH10B"

/note="Site: 1: SalI; Site: 2: BamHI; cDNA library was

prepared and sequenced in Mouse Genome Encyclopedia

Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in

FEATURES

source

RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGAGATTCGAGTCTTTTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 3.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCGAGTCTTTTAAATTAATACCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda E101 Cloning sites (5' end. Salt, 3' end. BamHI).

```

BASE COUNT      91 a    74 c    43 g    89 t
ORIGIN          100 b    60 d    10 e    1 f    1 g    1 h    1 i    1 j    1 k    1 l    1 m    1 n    1 o    1 p    1 q    1 r    1 s    1 t    1 u    1 v    1 w    1 x    1 y    1 z    1 aa    1 ab    1 ac    1 ad    1 ae    1 af    1 ag    1 ah    1 ai    1 aj    1 ak    1 al    1 am    1 an    1 ao    1 ap    1 aqu    1 ar    1 as    1 at    1 au    1 av    1 aw    1 ax    1 ay    1 az    1 ba    1 bb    1 bc    1 bd    1 be    1 bf    1 bg    1 bh    1 bi    1 bj    1 bk    1 bl    1 bm    1 bn    1 bo    1 bp    1 bq    1 br    1 bs    1 bt    1 bu    1 bv    1 bw    1 bx    1 by    1 bz    1 ca    1 cb    1 cc    1 cd    1 ce    1 cf    1 cg    1 ch    1 ci    1 cj    1 ck    1 cl    1 cm    1 cn    1 co    1 cp    1 cq    1 cr    1 cs    1 ct    1 cu    1 cv    1 cw    1 cx    1 cy    1 cz    1 da    1 db    1 dc    1 dd    1 de    1 df    1 dg    1 dh    1 di    1 dj    1 dk    1 dl    1 dm    1 dn    1 do    1 dp    1 dq    1 dr    1 ds    1 dt    1 du    1 dv    1 dw    1 dx    1 dy    1 dz    1 ea    1 eb    1 ec    1 ed    1 ee    1 ef    1 eg    1 eh    1 ei    1 ej    1 ek    1 el    1 em    1 en    1 eo    1 ep    1 eq    1 er    1 es    1 et    1 eu    1 ev    1 ew    1 ex    1 ey    1 ez    1 fa    1 fb    1 fc    1 fd    1 fe    1 ff    1 fg    1 fh    1 fi    1 fj    1 fk    1 fl    1 fm    1 fn    1 fo    1 fp    1 fq    1 fr    1 fs    1 ft    1 fu    1 fv    1 fw    1 fx    1 fy    1 fz    1 ga    1 gb    1 gc    1 gd    1 ge    1 gf    1 gg    1 gh    1 gi    1 gj    1 gk    1 gl    1 gm    1 gn    1 go    1 gp    1 gq    1 gr    1 gs    1 gt    1 gu    1 gv    1 gw    1 gx    1 gy    1 gz    1 ha    1 hb    1 hc    1 hd    1 he    1 hf    1 hg    1 hh    1 hi    1 hj    1 hk    1 hl    1 hm    1 hn    1 ho    1 hp    1 hq    1 hr    1 hs    1 ht    1 hu    1 hv    1 hw    1 hx    1 hy    1 hz    1 ia    1 ib    1 ic    1 id    1 ie    1 if    1 ig    1 ih    1 ii    1 ij    1 ik    1 il    1 im    1 in    1 io    1 ip    1 iq    1 ir    1 is    1 it    1 iu    1 iv    1 iw    1 ix    1 iy    1 iz    1 ja    1 jb    1 jc    1 jd    1 je    1 jf    1 jg    1 jh    1 ji    1 jj    1 jk    1 jl    1 jm    1 jn    1 jo    1 jp    1 jq    1 jr    1 js    1 jt    1 ju    1 jv    1 jw    1 jx    1 jy    1 jz    1 ka    1 kb    1 kc    1 kd    1 ke    1 kf    1 kg    1 kh    1 ki    1 kj    1 kk    1 kl    1 km    1 kn    1 ko    1 kp    1 kq    1 kr    1 ks    1 kt    1 ku    1 kv    1 kw    1 kx    1 ky    1 kz    1 la    1 lb    1 lc    1 ld    1 le    1 lf    1 lg    1 lh    1 li    1 lj    1 lk    1 ll    1 lm    1 ln    1 lo    1 lp    1 lq    1 lr    1 ls    1 lt    1 lu    1 lv    1 lw    1 lx    1 ly    1 lz    1 ma    1 mb    1 mc    1 md    1 me    1 mf    1 mg    1 mh    1 mi    1 mj    1 mk    1 ml    1 mm    1 mn    1 mo    1 mp    1 mq    1 mr    1 ms    1 mt    1 mu    1 mv    1 mw    1 mx    1 my    1 mz    1 na    1 nb    1 nc    1 nd    1 ne    1 nf    1 ng    1 nh    1 ni    1 nj    1 nk    1 nl    1 nm    1 no    1 np    1 nq    1 nr    1 ns    1 nt    1 nu    1 nv    1 nw    1 nx    1 ny    1 nz    1 oa    1 ob    1 oc    1 od    1 oe    1 of    1 og    1 oh    1 oi    1 oj    1 ok    1 ol    1 om    1 on    1 oo    1 op    1 oq    1 or    1 os    1 ot    1 ou    1 ov    1 ow    1 ox    1 oy    1 oz    1 pa    1 pb    1 pc    1 pd    1 pe    1 pf    1 pg    1 ph    1 pi    1 pj    1 pk    1 pl    1 pm    1 pn    1 po    1 pp    1 pq    1 pr    1 ps    1 pt    1 pu    1 pv    1 pw    1 px    1 py    1 pz    1 qa    1 qb    1 qc    1 qd    1 qe    1 qf    1 qg    1 qh    1 qi    1 qj    1 qk    1 ql    1 qm    1 qn    1 qo    1 qp    1 qq    1 qr    1 qs    1 qt    1 qu    1 qv    1 qw    1 qx    1 qy    1 qz    1 ra    1 rb    1 rc    1 rd    1 re    1 rf    1 rg    1 rh    1 ri    1 rj    1 rk    1 rl    1 rm    1 rn    1 ro    1 rp    1 rq    1 rr    1 rs    1 rt    1 ru    1 rv    1 rw    1 rx    1 ry    1 rz    1 sa    1 sb    1 sc    1 sd    1 se    1 sf    1 sg    1 sh    1 si    1 sj    1 sk    1 sl    1 sm    1 sn    1 so    1 sp    1 sq    1 sr    1 ss    1 st    1 su    1 sv    1 sw    1 sx    1 sy    1 sz    1 ta    1 tb    1 tc    1 td    1 te    1 tf    1 tg    1 th    1 ti    1 tj    1 tk    1 tl    1 tm    1 tn    1 to    1 tp    1 tq    1 tr    1 ts    1 tt    1 tu    1 tv    1 tw    1 tx    1 ty    1 tz    1 ua    1 ub    1 uc    1 ud    1 ue    1 uf    1 ug    1 uh    1 ui    1 uj    1 uk    1 ul    1 um    1 un    1 uo    1 up    1 uq    1 ur    1 us    1 ut    1 uu    1 uv    1 uw    1 ux    1 uy    1 uz    1 va    1 vb    1 vc    1 vd    1 ve    1 vf    1 vg    1 vh    1 vi    1 vj    1 vk    1 vl    1 vm    1 vn    1 vo    1 vp    1 vq    1 vr    1 vs    1 vt    1 vu    1 vv    1 vw    1 vx    1 vy    1 vz    1 wa    1 wb    1 wc    1 wd    1 we    1 wf    1 wg    1 wh    1 wi    1 wj    1 wk    1 wl    1 wm    1 wn    1 wo    1 wp    1 wq    1 wr    1 ws    1 wt    1 wu    1 wv    1 ww    1 wx    1 wy    1 wz    1 xa    1 xb    1 xc    1 xd    1 xe    1 xf    1 xg    1 xh    1 xi    1 xj    1 xk    1 xl    1 xm    1 xn    1 xo    1 xp    1 xq    1 xr    1 xs    1 xt    1 xu    1 xv    1 xw    1 xx    1 xy    1 xz    1 ya    1 yb    1 yc    1 yd    1 ye    1 yf    1 yg    1 yh    1 yi    1 yj    1 yk    1 yl    1 ym    1 yn    1 yo    1 yp    1 yq    1 yr    1 ys    1 yt    1 yu    1 yv    1 yw    1 yx    1 yy    1 yz    1 za    1 zb    1 zc    1 zd    1 ze    1 zf    1 zg    1 zh    1 zi    1 zj    1 zk    1 zl    1 zm    1 zn    1 zo    1 zp    1 zq    1 zr    1 zs    1 zt    1 zu    1 zv    1 zw    1 zx    1 zy    1 zz    1 aa    1 ab    1 ac    1 ad    1 ae    1 af    1 ag    1 ah    1 ai    1 aj    1 ak    1 al    1 am    1 an    1 ao    1 ap    1 aqu    1 ar    1 as    1 at    1 au    1 av    1 aw    1 ax    1 ay    1 az    1 ba    1 bb    1 bc    1 bd    1 be    1 bf    1 bg    1 bh    1 bi    1 bj    1 bk    1 bl    1 bm    1 bn    1 bo    1 bp    1 bq    1 br    1 bs    1 bt    1 bu    1 bv    1 bw    1 bx    1 by    1 bz    1 ca    1 cb    1 cc    1 cd    1 ce    1 cf    1 cg    1 ch    1 ci    1 cj    1 ck    1 cl    1 cm    1 cn    1 co    1 cp    1 cq    1 cr    1 cs    1 ct    1 cu    1 cv    1 cw    1 cx    1 cy    1 cz    1 da    1 db    1 dc    1 dd    1 de    1 df    1 dg    1 dh    1 di    1 dj    1 dk    1 dl    1 dm    1 dn    1 do    1 dp    1 dq    1 dr    1 ds    1 dt    1 du    1 dv    1 dw    1 dx    1 dy    1 dz    1 ea    1 eb    1 ec    1 ed    1 ee    1 ef    1 eg    1 eh    1 ei    1 ej    1 ek    1 el    1 em    1 en    1 eo    1 ep    1 eq    1 er    1 es    1 et    1 eu    1 ev    1 ew    1 ex    1 ey    1 ez    1 fa    1 fb    1 fc    1 fd    1 fe    1 ff    1 fg    1 fh    1 fi    1 fj    1 fk    1 fl    1 fm    1 fn    1 fo    1 fp    1 fq    1 fr    1 fs    1 ft    1 fu    1 fv    1 fw    1 fx    1 fy    1 fz    1 ga    1 gb    1 gc    1 gd    1 ge    1 gf    1 gg    1 gh    1 gi    1 gj    1 gk    1 gl    1 gm    1 gn    1 go    1 gp    1 gq    1 gr    1 gs    1 gt    1 gu    1 gv    1 gw    1 gx    1 gy    1 gz    1 ha    1 hb    1 hc    1 hd    1 he    1 hf    1 hg    1 hh    1 hi    1 hj    1 hk    1 hl    1 hm    1 hn    1 ho    1 hp    1 hq    1 hr    1 hs    1 ht    1 hu    1 hv    1 hw    1 hx    1 hy    1 hz    1 ia    1 ib    1 ic    1 id    1 ie    1 if    1 ig    1 ih    1 ii    1 ij    1 ik    1 il    1 im    1 in    1 io    1 ip    1 iq    1 ir    1 is    1 it    1 iu    1 iv    1 iw    1 ix    1 iy    1 iz    1 ja    1 jb    1 jc    1 jd    1 je    1 jf    1 jg    1 jh    1 ji    1 jj    1 jk    1 jl    1 jm    1 jn    1 jo    1 jp    1 jq    1 jr    1 js    1 jt    1 ju    1 jv    1 jw    1 jx    1 jy    1 jz    1 ka    1 kb    1 kc    1 kd    1 ke    1 kf    1 kg    1 kh    1 ki    1 kj    1 kk    1 kl    1 km    1 kn    1 ko    1 kp    1 kq    1 kr    1 ks    1 kt    1 ku    1 kv    1 kw    1 kx    1 ky    1 kz    1 la    1 lb    1 lc    1 ld    1 le    1 lf    1 lg    1 lh    1 li    1 lj    1 lk
```

High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES

SOURCE

```

FEATURES
  source
    Location/Qualifiers
      1..303
        /organism="Mus musculus"
        /db_xref="taxon:10090"
        /clone="B020040F24"
        /clone_lib="RIKEN full-length enriched, 2 cells egg"
        /tissue_type="egg"
        /dev_stage="2 cells"
        /lab_host="DH10B"
        /note="Site_1: Salt; Site_2: BamHI; cDNA library was
        prepared and sequenced in Mouse Genome Encyclopedia
        Project of Genome Exploration Research Group in Riken
        Genomic Sciences Center and Genome Science Laboratory in
        RIKEN, Division of Experimental Animal Research in Riken
        contributed to prepare mouse tissues. 1st strand cDNA was
        primed with a primer [5',
        GAGAGAGAGAGATCCAGAGCTCTTTTTTTTTTTTTTTTNN 3'], cDNA was
        prepared by using trihalose thermo-activated reverse
        transcriptase and subsequently enriched for full-length by the
        cap-trapper. second strand cDNA was prepared with the
        primer adapter of sequence [5',
        GAGAGAGAGATTCGATCTATTAAATTAATCCCCCCCCCCCC 3']. cDNA
        was cleaved with XhoI and BamHI. Vector: a modified
        pBluescript KS(-) after bulk excision from Lambda FLC I."
      85 a 65 c 70 g 83 t

```

BASE COUNT	85 a	65 c	70 g	83 t
ORIGIN				

Query Match 85.6%; Score 15.4; DB 9; Length 303;
Best Local Similarity 94.1%; Pred. No. 1.3e+03;
Matches 16: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

FEATURES

```

source
1. .337
/organism="Mus musculus"
/strain="CD-1"
/db_xref="taxon:10090"
/clone="IMAGE:3657339"
/clone_lib="McCarrey Eddy spermatoocytes"
/sex="male"
/tissue_type="spermatoocytes, pooled from multiple mice"
/dev_stage="60 day"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: testis; Vector: pBluescript SK+ (Stratagene
); Site_1: XhoI; Site_2: EcoRI; cDNA oligo dr-primed
[5'-(GA)10-CTAGTCTCGAGTTTTTTTTT-3'] and directionally
cloned using 5' linkers 5'-AATTCGCGCAG-3' and
5'-CTGCGCGC-3'. Size selection of >400bp material gives
average insert size ranging from 1-2 kb. Library was mass
excised (from lambda-UniZAP-XR) and resulting
single-stranded phagemids were prepped and transformed
into DH10B. Library contains 98% recombinants.
References: J. Androl. 20:635-639 and Gene 25:263-269.
Library constructed and donated by J. McCarrey, Ph.D.
(Southwest Foundation for Biomedical Research, Dept. of
Genetics); excision done by E.M. Eddy, Ph.D. (National
Institutes of Health, National Institute of Environmental
Health Sciences). Original lambda-based library is
available through ATCC, catalog #63422."
BASE COUNT      74 a      84 c      60 g      119 t
ORIGIN

```

```

Query Match      85.6%; Score 15.4; DB 10; Length 337;
Best Local Similarity 94.1%; Pred. No. 1.3e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 atactctggtcatgta 17
        |||||
Db      264 ATACTCTGTCATTTTA 280

```

Search completed: July 31, 2002, 12:12:16
Job time: 13489 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 14:11:25 ; Search time 720.8 Seconds
(without alignments)
42.875 Million cell updates/sec

Title: US-09-899-718A-8
Perfect score: 18
Sequence: 1 atactctggctcatgttaa 18

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*

- 1: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/NA1980.DAT.*
- 2: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/NA1981.DAT.*
- 3: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/NA1982.DAT.*
- 4: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/NA1983.DAT.*
- 5: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/NA1984.DAT.*
- 6: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/NA1985.DAT.*
- 7: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/NA1986.DAT.*
- 8: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/NA1987.DAT.*
- 9: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/NA1988.DAT.*
- 10: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/NA1989.DAT.*
- 11: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/NA1990.DAT.*
- 12: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/NA1991.DAT.*
- 13: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/NA1992.DAT.*
- 14: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/NA1993.DAT.*
- 15: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/NA1994.DAT.*
- 16: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/NA1995.DAT.*
- 17: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/NA1996.DAT.*
- 18: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/NA1997.DAT.*
- 19: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/NA1998.DAT.*
- 20: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/NA1999.DAT.*
- 21: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/NA2000.DAT.*
- 22: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/NA2001A.DAT.*
- 23: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/NA2001B.DAT.*
- 24: /SIDS1/gcgdata/hold-geneseq/geneqseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	16.4	91.1	32167	22	AAK90553 Human digestive sy
C 2	16.4	91.1	44848	21	AAK75080 Nucleotide sequenc
C 3	15.4	85.6	936	22	AAK68511 Human immune/haema
C 4	14.8	82.2	407	22	AAK57354 Human immune/haema
5	14.8	82.2	474	22	AAK66807 Human immune/haema
6	14.8	82.2	474	22	AAK66809 Human immune/haema
7	14.8	82.2	596	21	AAK10198 Fusarium venenatum
8	14.8	82.2	1327	21	AAK33932 Arabidopsis thalia
9	14.8	82.2	1339	21	AAK34572 Arabidopsis thalia

10	14.8	82.2	1382	21	AAK32880 Arabidopsis thalia
11	14.8	82.2	1554	21	AAK45097 Arabidopsis thalia
C 12	14.8	82.2	1926	23	ABL25967 Drosophila melanog
13	14.8	82.2	1935	20	AAZ30413 Arabidopsis thalia
14	14.8	82.2	1935	22	AAZ24587 Arabidopsis thalia
15	14.8	82.2	1943	22	AAK67740 Nucleotide sequenc
16	14.8	82.2	1943	22	AAK67743 Human immune/haema
17	14.8	82.2	2148	21	AAK44660 Arabidopsis thalia
C 18	14.8	82.2	2339	23	ABL16860 Arabidopsis thalia
19	14.8	82.2	2904	22	AAK70130 Drosophila melanog
20	14.8	82.2	3926	23	ABL25966 Human immune/haema
21	14.8	82.2	4104	23	ABL25964 Drosophila melanog
C 22	14.8	82.2	5320	22	AAK70131 Human immune/haema
23	14.8	82.2	15389	22	AAK80921 Human immune/haema
C 24	14.8	82.2	15610	22	AAK07072 Human reproductive
25	14.8	82.2	15610	22	AAK80920 Human immune/haema
26	14.8	82.2	39325	22	AAK81660 Human immune/haema
27	14.4	80.0	614	21	AAK08827 Human secreted pro
28	14.4	80.0	771	22	AAH07778 Human cDNA clone (
29	14.4	80.0	777	21	AAK02318 Human colon cancer
C 30	14.4	80.0	855	20	AAZ76142 ZOOA protein with
31	14.4	80.0	888	19	AAV32009 Human Rab protein
32	14.4	80.0	1408	19	AAV49254 Mouse kappa optate
33	14.4	80.0	1410	16	AAQ75926 Human cDNA sequenc
34	14.4	80.0	1551	22	AAH14958 Arabidopsis thalia
35	14.4	80.0	1717	21	AAK39158 Murine mGBP-3 CDNA
36	14.4	80.0	2134	23	AAK76101 Drosophila melanog
C 37	14.4	80.0	2486	21	AAZ89363 T. longibrachiatum
C 38	14.4	80.0	3345	23	ABL13550 Trichoderma cellob
39	14.4	80.0	4073	23	ABL08028 Human low adenosin
40	14.4	80.0	4123	17	AAQ91282 Human adenosine re
41	14.4	80.0	4124	21	AAK35057 Recombinant DNA fo
C 42	14.4	80.0	4124	21	AAK35057 Human low adenosin
C 43	14.4	80.0	4124	21	AAK35057 Human low adenosin
C 44	14.4	80.0	4729	11	AAQ04043 Human low adenosin
C 45	14.4	80.0	4740	21	AAK21154 Human low adenosin

ALIGNMENTS

RESULT 1
AAK90553/c
ID AAK90553 standard; DNA; 32167 BP.

XX AAK90553;

XX 05-NOV-2001 (first entry)

XX Human digestive system antigen genomic sequence SRQ ID NO: 4129.

DE Human digestive system antigen; gene therapy; cancer; appendicitis;

KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;

KW digestive system disorder; Meckel's diverticulum; ds.

XX Homo sapiens.

XX WO200155314-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01324.

XX 31-JAN-2000; 2000US-0179065.

XX 04-FEB-2000; 2000US-0180628.

XX 24-FEB-2000; 2000US-0184664.

XX 02-MAR-2000; 2000US-0186350.

XX 16-MAR-2000; 2000US-0189874.

XX 17-MAR-2000; 2000US-0190076.

XX 18-APR-2000; 2000US-0198123.

XX 07-MAY-2000; 2000US-0205515.

XX 28-JUN-2000; 2000US-0214886.

us-09-899-718a-8.rng

Thu Aug 1 08:30:07 2002

30-JUN-2000; 2000US-0215135.
07-JUL-2000; 2000US-0216647.
07-JUL-2000; 2000US-0216680.
11-JUL-2000; 2000US-0217487.
11-JUL-2000; 2000US-0217496.
14-JUL-2000; 2000US-0218290.
26-JUL-2000; 2000US-0220963.
26-JUL-2000; 2000US-0220964.
14-AUG-2000; 2000US-0224519.
14-AUG-2000; 2000US-0225213.
14-AUG-2000; 2000US-0225214.
14-AUG-2000; 2000US-0225266.
14-AUG-2000; 2000US-0225267.
14-AUG-2000; 2000US-0225268.
14-AUG-2000; 2000US-0225270.
14-AUG-2000; 2000US-0225447.
14-AUG-2000; 2000US-0225757.
14-AUG-2000; 2000US-0225758.
14-AUG-2000; 2000US-0225759.
18-AUG-2000; 2000US-0226279.
22-AUG-2000; 2000US-0226681.
22-AUG-2000; 2000US-0226686.
23-AUG-2000; 2000US-0227182.
23-AUG-2000; 2000US-0227009.
30-AUG-2000; 2000US-0228924.
01-SEP-2000; 2000US-0229287.
01-SEP-2000; 2000US-0229343.
01-SEP-2000; 2000US-0229344.
01-SEP-2000; 2000US-0229345.
05-SEP-2000; 2000US-0229509.
05-SEP-2000; 2000US-0229513.
06-SEP-2000; 2000US-0230437.
06-SEP-2000; 2000US-0230438.
08-SEP-2000; 2000US-0231242.
08-SEP-2000; 2000US-0231243.
08-SEP-2000; 2000US-0231244.
08-SEP-2000; 2000US-0231413.
08-SEP-2000; 2000US-0231414.
08-SEP-2000; 2000US-0232080.
08-SEP-2000; 2000US-0232081.
12-SEP-2000; 2000US-0231968.
14-SEP-2000; 2000US-0232397.
14-SEP-2000; 2000US-0232398.
14-SEP-2000; 2000US-0232399.
14-SEP-2000; 2000US-0232400.
14-SEP-2000; 2000US-0232401.
14-SEP-2000; 2000US-0233063.
14-SEP-2000; 2000US-0233064.
14-SEP-2000; 2000US-0233065.
21-SEP-2000; 2000US-0234223.
21-SEP-2000; 2000US-0234274.
25-SEP-2000; 2000US-0234997.
25-SEP-2000; 2000US-0234998.
26-SEP-2000; 2000US-0235484.
27-SEP-2000; 2000US-0235834.
27-SEP-2000; 2000US-0235836.
29-SEP-2000; 2000US-0236327.
29-SEP-2000; 2000US-0236367.
29-SEP-2000; 2000US-0236368.
29-SEP-2000; 2000US-0236369.
29-SEP-2000; 2000US-0236370.
02-OCT-2000; 2000US-0236802.
02-OCT-2000; 2000US-0237037.
02-OCT-2000; 2000US-0237038.
02-OCT-2000; 2000US-0237039.
02-OCT-2000; 2000US-0237040.
13-OCT-2000; 2000US-0239935.
13-OCT-2000; 2000US-0239937.
20-OCT-2000; 2000US-0240960.
20-OCT-2000; 2000US-0241221.
20-OCT-2000; 2000US-0241785.
20-OCT-2000; 2000US-0241786.
20-OCT-2000; 2000US-0241787.
20-OCT-2000; 2000US-0241808.
20-OCT-2000; 2000US-0241809.
20-OCT-2000; 2000US-0241826.
01-NOV-2000; 2000US-0244617.
08-NOV-2000; 2000US-0246474.
08-NOV-2000; 2000US-0246475.
08-NOV-2000; 2000US-0246476.
08-NOV-2000; 2000US-0246477.
08-NOV-2000; 2000US-0246478.
08-NOV-2000; 2000US-0246523.
08-NOV-2000; 2000US-0246524.
08-NOV-2000; 2000US-0246525.
08-NOV-2000; 2000US-0246526.
08-NOV-2000; 2000US-0246527.
08-NOV-2000; 2000US-0246528.
08-NOV-2000; 2000US-0246532.
08-NOV-2000; 2000US-0246609.
08-NOV-2000; 2000US-0246610.
08-NOV-2000; 2000US-0246611.
08-NOV-2000; 2000US-0246613.
17-NOV-2000; 2000US-0249207.
17-NOV-2000; 2000US-0249208.
17-NOV-2000; 2000US-0249210.
17-NOV-2000; 2000US-0249211.
17-NOV-2000; 2000US-0249212.
17-NOV-2000; 2000US-0249213.
17-NOV-2000; 2000US-0249214.
17-NOV-2000; 2000US-0249215.
17-NOV-2000; 2000US-0249216.
17-NOV-2000; 2000US-0249217.
17-NOV-2000; 2000US-0249218.
17-NOV-2000; 2000US-0249244.
17-NOV-2000; 2000US-0249245.
17-NOV-2000; 2000US-0249264.
17-NOV-2000; 2000US-0249265.
17-NOV-2000; 2000US-0249297.
17-NOV-2000; 2000US-0249299.
17-NOV-2000; 2000US-0249300.
01-DEC-2000; 2000US-0250160.
05-DEC-2000; 2000US-0250391.
05-DEC-2000; 2000US-0251030.
05-DEC-2000; 2000US-0251988.
05-DEC-2000; 2000US-0256719.
06-DEC-2000; 2000US-0251479.
08-DEC-2000; 2000US-0251856.
08-DEC-2000; 2000US-0251868.
08-DEC-2000; 2000US-0251869.
08-DEC-2000; 2000US-0251989.
08-DEC-2000; 2000US-0251990.
11-DEC-2000; 2000US-0254097.
05-JAN-2001; 2001US-0259678.
(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-502630/55.
Polynucleotides encoding digestive system antigens, useful for
diagnosing, treating, preventing and/or prognosing disorders of the
digestive system, particularly cancer and cancer metastases -
Disclosure; SEQ ID NO 4129; 986pp; English.
The present invention provides the protein and coding sequences of a
number of human digestive system antigens. These can be used in the
diagnosis, treatment and prevention of digestive system disorders,
including cancer, Meckel's diverticulum, bacterial or parasitic
infections, appendicitis, Hirschsprung's disease, chronic colitis or
ulcerative colitis. The present sequence is a genomic DNA fragment
encoding a digestive system antigen of the invention.

XX SQ Sequence 32167 BP; 9824 A; 5983 C; 6230 G; 10130 T; 0 other;

Query Match 91.1%; Score 16.4; DB 22; Length 32167;
Best Local Similarity 94.4%; Pred. No. 53;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atactctggtcatgttaa 18
||| ||||| ||||| |||||
Db 31080 ATGCTCTGGTCATGTAA 31063

RESULT 2
AAA75080/c
ID AAA75080 standard; DNA; 44848 BP.

XX AC AAA75080;

XX DT 15-JAN-2001 (first entry)

XX DE Nucleotide sequence of the human heparanase gene.

XX KW Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;
KW heparin-binding growth factor; cytokine; neurodegenerative plaque;
KW wound healing; infection; burn; angiogenesis; restenosis;
KW atherosclerosis; inflammation; neurodegenerative disease;
KW Gerstmann-Straussler Syndrome; Creutzfeldt-Jakob disease; ds.

XX OS Homo sapiens.

XX PN WO200052178-A1.

XX PD 08-SEP-2000.

XX PF 14-FEB-2000; 2000WO-US03542.

XX PR 01-MAR-1999; 99US-0258892.

XX PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
XX PA (FRIE/) FRIEDMAN M M.

XX PI Pecker I, Vlodavsky I, Feinstein E;

XX DR WPI; 2000-579289/54.

XX PT New polynucleotides encoding a polypeptide having heparanase activity,
XX PT useful in wound healing and in gene therapy, particularly in treating
XX PT tumour, inflammation, autoimmunity, neurodegenerative diseases -

XX PS Claim 9; Page 131-143; 152pp; English.

XX CC The present sequence represents a human gene which encodes a protein
XX CC with heparanase catalytic activity. The heparanase (hpa) polynucleotide
XX CC is useful in gene therapy, particularly in treating tumour, inflammation
XX CC or autoimmunity. Particularly, the polynucleotide is useful in modulating
XX CC the bioavailability of heparin-binding growth factors, cellular responses
XX CC to heparin-binding growth factors (e.g. bFGF) and cytokines
XX CC (e.g. interleukin (IL)-8), cell interaction with plasma lipoproteins,
XX CC cellular susceptibility to certain viral and some bacterial and protozoa
XX CC infections, or disintegration of neurodegenerative plaques. The
XX CC polynucleotide is also useful in wound healing (e.g. thermal, chemical
XX CC or radiation burns), and in the treatment of angiogenesis, restenosis,
XX CC atherosclerosis, inflammation, neurodegenerative diseases (Gerstmann-
XX CC Straussler Syndrome or Creutzfeldt-Jakob disease), and some viral,
XX CC bacterial or protozoa infections.

XX SQ Sequence 44848 BP; 12560 A; 9646 C; 8930 G; 13712 T; 0 other;

Query Match 91.1%; Score 16.4; DB 21; Length 44848;
Best Local Similarity 94.4%; Pred. No. 55;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atactctggtcatgttaa 18
| ||||| ||||| |||||
Db 11410 AGACTCTGGTCATGTAA 11393

RESULT 3
AAK68511/c
ID AAK68511 standard; DNA; 936 BP.

XX AC AAK68511;

XX DT 06-NOV-2001 (first entry)

XX DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23323.

XX KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX KW cytostatic; gene therapy; vaccine; metastasis; ds.

XX OS Homo sapiens.

XX PN WO200157182-A2.

XX PD 09-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01354.

XX PR 31-JAN-2000; 2000US-0179065.

XX PR 04-FEB-2000; 2000US-0180628.

XX PR 24-FEB-2000; 2000US-0184664.

XX PR 02-MAR-2000; 2000US-0186350.

XX PR 16-MAR-2000; 2000US-0189874.

XX PR 17-MAR-2000; 2000US-0190076.

XX PR 18-APR-2000; 2000US-0198123.

XX PR 19-MAY-2000; 2000US-0205515.

XX PR 07-JUN-2000; 2000US-0209467.

XX PR 28-JUN-2000; 2000US-0214886.

XX PR 30-JUN-2000; 2000US-0215135.

XX PR 07-JUL-2000; 2000US-0216647.

XX PR 07-JUL-2000; 2000US-0216880.

XX PR 11-JUL-2000; 2000US-0217487.

XX PR 11-JUL-2000; 2000US-0217496.

XX PR 14-JUL-2000; 2000US-0218290.

XX PR 26-JUL-2000; 2000US-0220963.

XX PR 26-JUL-2000; 2000US-0220964.

XX PR 14-AUG-2000; 2000US-0224518.

XX PR 14-AUG-2000; 2000US-0224519.

XX PR 14-AUG-2000; 2000US-0225213.

XX PR 14-AUG-2000; 2000US-0225214.

XX PR 14-AUG-2000; 2000US-0225266.

XX PR 14-AUG-2000; 2000US-0225267.

XX PR 14-AUG-2000; 2000US-0225268.

XX PR 14-AUG-2000; 2000US-0225270.

XX PR 14-AUG-2000; 2000US-0225447.

XX PR 14-AUG-2000; 2000US-0225757.

XX PR 14-AUG-2000; 2000US-0225758.

XX PR 14-AUG-2000; 2000US-0225759.

XX PR 18-AUG-2000; 2000US-0226279.

XX PR 22-AUG-2000; 2000US-0226681.

XX PR 22-AUG-2000; 2000US-0226868.

XX PR 23-AUG-2000; 2000US-0227182.

XX PR 23-AUG-2000; 2000US-0227009.

XX PR 30-AUG-2000; 2000US-0228924.

XX PR 01-SEP-2000; 2000US-0229287.

XX PR 01-SEP-2000; 2000US-0229343.

XX PR 01-SEP-2000; 2000US-0229344.

XX PR 05-SEP-2000; 2000US-0229345.

XX PR 05-SEP-2000; 2000US-0229509.

XX PR 05-SEP-2000; 2000US-0229513.

XX PR 06-SEP-2000; 2000US-0230437.

XX PR 06-SEP-2000; 2000US-0230438.

XX PR 08-SEP-2000; 2000US-0231242.

XX	Homo sapiens.	PR	26-SEP-2000;	2000US-0235484.
OS		PR	27-SEP-2000;	2000US-0235834.
XX		PR	27-SEP-2000;	2000US-0235836.
PN	WO200157182-A2.	PR	29-SEP-2000;	2000US-0236327.
XX		PR	29-SEP-2000;	2000US-0236367.
PD		PR	29-SEP-2000;	2000US-0236368.
XX	09-AUG-2001.	PR	29-SEP-2000;	2000US-0236369.
XX		PR	29-SEP-2000;	2000US-0236370.
PF		PR	02-OCT-2000;	2000US-0236802.
XX	17-JAN-2001; 2001WO-US01354.	PR	02-OCT-2000;	2000US-0237037.
XX		PR	02-OCT-2000;	2000US-0237038.
PR	31-JAN-2000; 2000US-0179065.	PR	02-OCT-2000;	2000US-0237039.
PR	04-FEB-2000; 2000US-0180628.	PR	02-OCT-2000;	2000US-0237040.
PR	24-FEB-2000; 2000US-0184664.	PR	13-OCT-2000;	2000US-0239935.
PR	02-MAR-2000; 2000US-0186350.	PR	13-OCT-2000;	2000US-0239937.
PR	16-MAR-2000; 2000US-0189874.	PR	20-OCT-2000;	2000US-0240960.
PR	17-MAR-2000; 2000US-0190076.	PR	20-OCT-2000;	2000US-0241221.
PR	18-APR-2000; 2000US-0198123.	PR	20-OCT-2000;	2000US-0241785.
PR	19-MAY-2000; 2000US-0205515.	PR	20-OCT-2000;	2000US-0241786.
PR	07-JUN-2000; 2000US-0209467.	PR	20-OCT-2000;	2000US-0241787.
PR	28-JUN-2000; 2000US-0214886.	PR	20-OCT-2000;	2000US-0241808.
PR	30-JUN-2000; 2000US-0215135.	PR	20-OCT-2000;	2000US-0241809.
PR	07-JUL-2000; 2000US-0216647.	PR	20-OCT-2000;	2000US-0241826.
PR	07-JUL-2000; 2000US-0216880.	PR	01-NOV-2000;	2000US-0244617.
PR	11-JUL-2000; 2000US-0217487.	PR	08-NOV-2000;	2000US-0246474.
PR	11-JUL-2000; 2000US-0217496.	PR	08-NOV-2000;	2000US-0246475.
PR	14-JUL-2000; 2000US-0218290.	PR	08-NOV-2000;	2000US-0246476.
PR	26-JUL-2000; 2000US-0220963.	PR	08-NOV-2000;	2000US-0246477.
PR	26-JUL-2000; 2000US-0220964.	PR	08-NOV-2000;	2000US-0246478.
PR	14-AUG-2000; 2000US-0224518.	PR	08-NOV-2000;	2000US-0246523.
PR	14-AUG-2000; 2000US-0224519.	PR	08-NOV-2000;	2000US-0246524.
PR	14-AUG-2000; 2000US-0225213.	PR	08-NOV-2000;	2000US-0246525.
PR	14-AUG-2000; 2000US-0225214.	PR	08-NOV-2000;	2000US-0246526.
PR	14-AUG-2000; 2000US-0225266.	PR	08-NOV-2000;	2000US-0246527.
PR	14-AUG-2000; 2000US-0225267.	PR	08-NOV-2000;	2000US-0246528.
PR	14-AUG-2000; 2000US-0225268.	PR	08-NOV-2000;	2000US-0246532.
PR	14-AUG-2000; 2000US-0225447.	PR	08-NOV-2000;	2000US-0246609.
PR	14-AUG-2000; 2000US-0225757.	PR	08-NOV-2000;	2000US-0246610.
PR	14-AUG-2000; 2000US-0225758.	PR	08-NOV-2000;	2000US-0246611.
PR	14-AUG-2000; 2000US-0225759.	PR	08-NOV-2000;	2000US-0246613.
PR	18-AUG-2000; 2000US-0226279.	PR	17-NOV-2000;	2000US-0249207.
PR	22-AUG-2000; 2000US-0226681.	PR	17-NOV-2000;	2000US-0249208.
PR	22-AUG-2000; 2000US-0226688.	PR	17-NOV-2000;	2000US-0249209.
PR	22-AUG-2000; 2000US-0227182.	PR	17-NOV-2000;	2000US-0249210.
PR	23-AUG-2000; 2000US-0227182.	PR	17-NOV-2000;	2000US-0249211.
PR	30-AUG-2000; 2000US-0228924.	PR	17-NOV-2000;	2000US-0249212.
PR	01-SEP-2000; 2000US-0229287.	PR	17-NOV-2000;	2000US-0249213.
PR	01-SEP-2000; 2000US-0229343.	PR	17-NOV-2000;	2000US-0249214.
PR	01-SEP-2000; 2000US-0229344.	PR	17-NOV-2000;	2000US-0249215.
PR	01-SEP-2000; 2000US-0229345.	PR	17-NOV-2000;	2000US-0249216.
PR	05-SEP-2000; 2000US-0229509.	PR	17-NOV-2000;	2000US-0249217.
PR	05-SEP-2000; 2000US-0229513.	PR	17-NOV-2000;	2000US-0249218.
PR	06-SEP-2000; 2000US-0230437.	PR	17-NOV-2000;	2000US-0249244.
PR	06-SEP-2000; 2000US-0230438.	PR	17-NOV-2000;	2000US-0249245.
PR	08-SEP-2000; 2000US-0231242.	PR	17-NOV-2000;	2000US-0249264.
PR	08-SEP-2000; 2000US-0231243.	PR	17-NOV-2000;	2000US-0249265.
PR	08-SEP-2000; 2000US-0231244.	PR	17-NOV-2000;	2000US-0249297.
PR	08-SEP-2000; 2000US-0231413.	PR	17-NOV-2000;	2000US-0249299.
PR	08-SEP-2000; 2000US-0231414.	PR	17-NOV-2000;	2000US-0249300.
PR	08-SEP-2000; 2000US-0232080.	PR	01-DEC-2000;	2000US-0250160.
PR	08-SEP-2000; 2000US-0232081.	PR	01-DEC-2000;	2000US-0250391.
PR	12-SEP-2000; 2000US-0231968.	PR	05-DEC-2000;	2000US-0251030.
PR	14-SEP-2000; 2000US-0232397.	PR	05-DEC-2000;	2000US-0251988.
PR	14-SEP-2000; 2000US-0232398.	PR	05-DEC-2000;	2000US-0251989.
PR	14-SEP-2000; 2000US-0232399.	PR	06-DEC-2000;	2000US-0251990.
PR	14-SEP-2000; 2000US-0232400.	PR	06-DEC-2000;	2000US-0251997.
PR	14-SEP-2000; 2000US-0232401.	PR	08-DEC-2000;	2000US-0251998.
PR	14-SEP-2000; 2000US-0233063.	PR	08-DEC-2000;	2000US-0251999.
PR	14-SEP-2000; 2000US-0233064.	PR	11-DEC-2000;	2000US-0254097.
PR	21-SEP-2000; 2000US-0233065.	PR	05-JAN-2001;	2001US-0259678.
PR	21-SEP-2000; 2000US-0234223.	PR		
PR	21-SEP-2000; 2000US-0234274.	PR		
PR	25-SEP-2000; 2000US-0234997.	PR		
PR	25-SEP-2000; 2000US-0234998.	XX		
PR		XX		

PA (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Barash SC, Ruben SM;
 PI
 XX WPI; 2001-483426/52.
 DR P-PSDB; AAK84573.
 XX
 PT Nucleic acids encoding human immune/haematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and
 PT metastasis -
 XX
 XX Claim 1; SEQ ID NO 2414; 3071pp + Sequence Listing; English.
 PS
 XX
 CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
 CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patient's own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/haematopoietic-related diseases, especially
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/haematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
 CC represent sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 407 BP; 105 A; 102 C; 56 G; 140 T; 4 other;

Query Match 82.2%; Score 14.8; DB 22; Length 407;
 Best Local Similarity 88.9%; Pred. No. 2.4e+02;
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atactctggctatgtaa 18
 ||| ||||| |||||
 Db 133 ATAGTCTGGTAATGTTAA 116

RESULT 5
 AAK66807
 ID AAK66807 standard; DNA; 474 BP.
 XX
 AC AAK66807;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:21619.
 XX
 KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
 KW cytostatic; gene therapy; vaccine; metastasis; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200157182-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01354.
 XX
 XX 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.

PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPT: 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides, useful for preventing, diagnosing and/or treating cancers and metastasis -

Disclosure: SEQ ID NO 21619; 307lpp + Sequence Listing: English.

AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I) amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic activity, and can be used in gene therapy and vaccine production. (I) proteins and polynucleotides may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate (I) expression. For example, they may be used to treat disorders associated with decreased

CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 474 BP; 135 A; 109 C; 90 G; 140 T; 0 other;

Query Match 82.2%; Score 14.8; DB 22; Length 474;

Best Local Similarity 88.9%; Pred. No. 2.5e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 atactctgtcatgttaa 18
||| ||||| |||||
Db 392 ataactgttcaagttaa 409

RESULT 6

AAK66809
ID AAK66809 standard; DNA; 474 BP.

XX AAK66809;

AC AAK66809;

XX 06-NOV-2001 (first entry)

DE Human immune/hematopoietic antigen genomic sequence SEQ ID NO:21621.
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.

XX Homo sapiens.

XX WO200157182-A2.

XX 09-AUG-2001.

PF 17-JAN-2001; 2001WO-US01354.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 02-FEB-2000; 2000US-0184664.

PR 16-MAR-2000; 2000US-0186350.

PR 17-MAR-2000; 2000US-0189874.

PR 18-APR-2000; 2000US-0190076.

PR 19-MAY-2000; 2000US-0198123.

PR 07-JUN-2000; 2000US-0205515.

PR 28-JUN-2000; 2000US-0209467.

PR 30-JUN-2000; 2000US-0214886.

PR 07-JUL-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

st Local Similarity	88.9%;	Pred. No.	2.5e+02;
Mismatches	0:	Mismatches	2:
Indels	0:	Gaps	0:

Best Local Similarity	88.9%;	Pred. No. 2.5e+02;	Indels 0
Matches	16;	Conservative 0;	Mismatches 2;

QY 1 atactctggtcatgttaa 18
 Db 392 ataactggtcaagttaa 409

RESULT 7
 ID AAF10198
 AC AAF10198;
 DT 13-MAR-2001 (first entry)
 DE Fusarium venenatum EST SEQ ID NO:2721.

XX Multiple gene expression; filamentous fungal cell; EST;
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.
 XX Fusarium venenatum.

XX WO2000056762-A2.
 PN 28-SEP-2000.
 XX 22-MAR-2000; 2000WO-US07781.
 XX 22-MAR-1999; 99US-0273623.

XX (NOVO) NOVO NORDISK BIOTECH INC.
 PA (NOVO) NOVO NORDISK AS.

XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
 XX WPI; 2000-594572/56.

XX Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags -

XX Claim 86; Page 1372-1373; 3161pp; English.

XX The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an
 CC array equals one gene or open reading frame, and organisation of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
 CC Fusarium venenatum; AAF11248 to AAF11953 represents ESTs from Aspergillus
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
 CC all specifically claimed in the present invention.

XX Sequence 596 BP; 156 A; 146 C; 149 G; 139 T; 6 other;

Query Match

82.2%; Score 14.8; DB 21; Length 596;

Best Local Similarity 88.9%; Pred. No. 2.5e+02;
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atactctggtcatgttaa 18
 Db 565 atactctggtcatgttaa 582

RESULT 8
 ID AAC33932
 AC AAC33932;
 DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 4839.
 DE Arabidopsis thaliana
 XX Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway;
 KW metabolic pathway; promoter; termination sequence; ss.
 XX Arabidopsis thaliana.

XX EF1033405-A2.
 PN 06-SEP-2000.
 XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139452.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 28-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145270.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0145386.
 PR 02-AUG-1999; 99US-0145388.
 PR 02-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.

PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 82.2%; Score 14.8; DB 21; Length 1327;
 Best Local Similarity 88.9%; Pred. No. 2.7e+02;
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 atactctgttcattgttaa 18
 || ||||| |||||
 Db 269 atccctgtgttattgttaa 286

RESULT 9
AAC34572
ID AAC34572 standard; DNA; 1339 BP.
XX AC AAC34572;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 7122.
XX KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126284.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 24-MAY-1999; 99US-0135353.
PR 25-MAY-1999; 99US-0135629.
PR 27-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136382.
PR 01-JUN-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137502.
PR 08-JUN-1999; 99US-0137724.
PR 10-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 20-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.

us-09-899-718a-8.rng

Thu Aug 1 08:30:07 2002

PR 20-AUG-1999;	99US-0149723.	DT 17-OCT-2000 (first entry)	99US-0121825.
PR 20-AUG-1999;	99US-0149929.	XX Arabidopsis thaliana DNA fragment SEQ ID NO: 983.	99US-0123180.
PR 23-AUG-1999;	99US-0149902.	XX Hybridisation assay; genetic mapping; gene expression control;	99US-0123548.
PR 23-AUG-1999;	99US-0149930.	KW protein identification; signal transduction pathway;	99US-0125788.
PR 25-AUG-1999;	99US-0150566.	KW metabolic pathway; promoter; termination sequence; ss.	99US-0126264.
PR 26-AUG-1999;	99US-0150884.	XX Arabidopsis thaliana.	99US-0126785.
PR 27-AUG-1999;	99US-0151065.	OS	99US-0127462.
PR 27-AUG-1999;	99US-0151066.	XX	99US-0128234.
PR 27-AUG-1999;	99US-0151080.	XX	99US-0128714.
PR 30-AUG-1999;	99US-0151303.	PN	99US-0128845.
PR 31-AUG-1999;	99US-0151438.	XX	99US-0130077.
PR 01-SEP-1999;	99US-0151930.	PD	99US-0130449.
PR 07-SEP-1999;	99US-0152363.	XX	99US-0130510.
PR 10-SEP-1999;	99US-0153070.	PF	99US-0130891.
PR 13-SEP-1999;	99US-0153758.	XX	99US-0131449.
PR 15-SEP-1999;	99US-0154018.	PR	99US-0132048.
PR 16-SEP-1999;	99US-0154039.	PR	99US-0132407.
PR 20-SEP-1999;	99US-0154779.	PR	99US-0132484.
PR 22-SEP-1999;	99US-0155139.	PR	99US-0132485.
PR 23-SEP-1999;	99US-0155486.	PR	99US-0132486.
PR 24-SEP-1999;	99US-0155659.	PR	99US-0132487.
PR 28-SEP-1999;	99US-0156458.	PR	99US-0132863.
PR 29-SEP-1999;	99US-0156596.	PR	99US-0134256.
PR 04-OCT-1999;	99US-0157117.	PR	99US-0134218.
PR 05-OCT-1999;	99US-0157753.	PR	99US-0134219.
PR 06-OCT-1999;	99US-0157865.	PR	99US-0134221.
PR 07-OCT-1999;	99US-0158029.	PR	99US-0134370.
PR 08-OCT-1999;	99US-0158222.	PR	99US-0134768.
PR 12-OCT-1999;	99US-0158369.	PR	99US-0134941.
PR 13-OCT-1999;	99US-0159293.	PR	99US-0135124.
PR 13-OCT-1999;	99US-0159294.	PR	99US-0135353.
PR 13-OCT-1999;	99US-0159295.	PR	99US-0135629.
PR 14-OCT-1999;	99US-0159329.	PR	99US-0136021.
PR 14-OCT-1999;	99US-0159330.	PR	99US-0136392.
PR 14-OCT-1999;	99US-0159331.	PR	99US-0136782.
PR 14-OCT-1999;	99US-0159637.	PR	99US-0137222.
PR 14-OCT-1999;	99US-0159638.	PR	99US-0137528.
PR 14-OCT-1999;	99US-0159584.	PR	99US-0137724.
PR 21-OCT-1999;	99US-0160767.	PR	99US-0138094.
PR 21-OCT-1999;	99US-0160768.	PR	99US-0138540.
PR 21-OCT-1999;	99US-0160770.	PR	99US-0138847.
PR 21-OCT-1999;	99US-0160814.	PR	99US-0139119.
PR 21-OCT-1999;	99US-0160815.	PR	99US-0139452.
PR 22-OCT-1999;	99US-0160981.	PR	99US-0139453.
PR 22-OCT-1999;	99US-0160989.	PR	99US-0139492.
PR 25-OCT-1999;	99US-0161404.	PR	99US-0139494.
PR 25-OCT-1999;	99US-0161405.	PR	99US-0139454.
PR 25-OCT-1999;	99US-0161406.	PR	99US-0139455.
PR 26-OCT-1999;	99US-0161359.	PR	99US-0139456.
PR 26-OCT-1999;	99US-0161360.	PR	99US-0139457.
PR 26-OCT-1999;	99US-0161361.	PR	99US-0139458.
PR 28-OCT-1999;	99US-0161920.	PR	99US-0139459.
PR 28-OCT-1999;	99US-0161992.	PR	99US-0139460.
PR 28-OCT-1999;	99US-0161993.	PR	99US-0139461.
PR 28-OCT-1999;	99US-0161994.	PR	99US-0139462.
PR 29-OCT-1999;	99US-0162142.	PR	99US-0139463.
PR 29-OCT-1999;		PR	99US-0139750.

Query Match	82.28;	Score 14.8;	DB 21;	Length 1339;
Best Local Similarity	88.98;	Pred. No. 2.7e+02;		
Matches 16;	Conservative	0;	Mismatches 2;	Indels 0;
Gaps	0;			

QY	1	atactctggtcatgtaa	18
DB	264	atcctctggtcatgtaa	281

RESULT	10
AAC32880	
ID AAC32880 standard; DNA; 1382 BP.	
XX	
AC AAC32880;	
XX	

PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139859.
 PR 23-JUN-1999; 99US-0140353.
 PR 24-JUN-1999; 99US-0140354.
 PR 28-JUN-1999; 99US-0140695.
 PR 29-JUN-1999; 99US-0140823.
 PR 30-JUN-1999; 99US-0140991.
 PR 01-JUL-1999; 99US-0141287.
 PR 02-JUL-1999; 99US-0141842.
 PR 06-JUL-1999; 99US-0142055.
 PR 08-JUL-1999; 99US-0142390.
 PR 09-JUL-1999; 99US-0142803.
 PR 12-JUL-1999; 99US-0142920.
 PR 13-JUL-1999; 99US-0142977.
 PR 14-JUL-1999; 99US-0143542.
 PR 15-JUL-1999; 99US-0143624.
 PR 16-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 19-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 20-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 21-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 22-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 23-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 26-JUL-1999; 99US-0145224.
 PR 27-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 28-JUL-1999; 99US-0145919.
 PR 02-AUG-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 04-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 05-AUG-1999; 99US-0147304.
 PR 05-AUG-1999; 99US-0147192.
 PR 06-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 09-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0148684.
 PR 17-AUG-1999; 99US-0149368.
 PR 18-AUG-1999; 99US-0149175.
 PR 20-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 23-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149929.
 PR 25-AUG-1999; 99US-0149930.
 PR 26-AUG-1999; 99US-0150566.
 PR 27-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.

PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155139.
 PR 24-SEP-1999; 99US-0155486.
 PR 28-SEP-1999; 99US-0155659.
 PR 29-SEP-1999; 99US-0156458.
 PR 04-OCT-1999; 99US-0156596.
 PR 05-OCT-1999; 99US-0157117.
 PR 06-OCT-1999; 99US-0157753.
 PR 07-OCT-1999; 99US-0157865.
 PR 08-OCT-1999; 99US-0158029.
 PR 12-OCT-1999; 99US-0158232.
 PR 13-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 14-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 22-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 25-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 26-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161922.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 82.2%; Score 14.8; DB 21; Length 1382;
 Best Local Similarity 88.9%; Pred. No. 2.7e+02;
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 atactctggctatgtaa 18
 || ||||| |||||
 Db 349 atcctctggctatgtaa 366

RESULT 11
 AAC45097
 ID AAC45097 standard; DNA; 1554 BP.
 XX
 AC AAC45097;
 XX
 DT 18-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana DNA fragment SEQ ID NO: 45288.
 XX
 KW Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway;
 KW metabolic pathway; promoter; termination sequence; ss.

us-09-899-718a-8.rng

Thu Aug 1 08:30:07 2002

XX OS Arabidopsis thaliana.	PR 29-JUN-1999;	99US-0140991.
XX PN EP1033405-A2.	PR 30-JUN-1999;	99US-0141287.
XX PD 06-SEP-2000.	PR 01-JUL-1999;	99US-0141842.
XX PF 25-FEB-2000; 2000EP-0301439.	PR 02-JUL-1999;	99US-0142154.
XX XX 25-FEB-1999;	PR 03-JUL-1999;	99US-0142390.
XX XX 05-MAR-1999;	PR 04-JUL-1999;	99US-0142803.
XX XX 09-MAR-1999;	PR 05-JUL-1999;	99US-0142920.
XX XX 23-MAR-1999;	PR 06-JUL-1999;	99US-0143542.
XX XX 25-MAR-1999;	PR 07-JUL-1999;	99US-0143624.
XX XX 29-MAR-1999;	PR 08-JUL-1999;	99US-0144005.
XX XX 01-APR-1999;	PR 09-JUL-1999;	99US-0144085.
XX XX 06-APR-1999;	PR 10-JUL-1999;	99US-0144086.
XX XX 08-APR-1999;	PR 11-JUL-1999;	99US-0144325.
XX XX 16-APR-1999;	PR 12-JUL-1999;	99US-0144331.
XX XX 19-APR-1999;	PR 13-JUL-1999;	99US-0144332.
XX XX 21-APR-1999;	PR 14-JUL-1999;	99US-0144333.
XX XX 23-APR-1999;	PR 15-JUL-1999;	99US-0144334.
XX XX 28-APR-1999;	PR 16-JUL-1999;	99US-0144335.
XX XX 30-APR-1999;	PR 17-JUL-1999;	99US-0144352.
XX XX 30-APR-1999;	PR 18-JUL-1999;	99US-0144632.
XX XX 04-MAY-1999;	PR 19-JUL-1999;	99US-0144884.
XX XX 05-MAY-1999;	PR 20-JUL-1999;	99US-0145086.
XX XX 06-MAY-1999;	PR 21-JUL-1999;	99US-0145088.
XX XX 07-MAY-1999;	PR 22-JUL-1999;	99US-0145087.
XX XX 11-MAY-1999;	PR 23-JUL-1999;	99US-0145089.
XX XX 14-MAY-1999;	PR 24-JUL-1999;	99US-0145192.
XX XX 14-MAY-1999;	PR 25-JUL-1999;	99US-0145145.
XX XX 14-MAY-1999;	PR 26-JUL-1999;	99US-0145218.
XX XX 14-MAY-1999;	PR 27-JUL-1999;	99US-0145224.
XX XX 14-MAY-1999;	PR 28-JUL-1999;	99US-0145276.
XX XX 14-MAY-1999;	PR 29-JUL-1999;	99US-0145913.
XX XX 14-MAY-1999;	PR 30-JUL-1999;	99US-0145918.
XX XX 14-MAY-1999;	PR 31-JUL-1999;	99US-0145919.
XX XX 14-MAY-1999;	PR 01-AUG-1999;	99US-0145951.
XX XX 14-MAY-1999;	PR 02-AUG-1999;	99US-0146386.
XX XX 14-MAY-1999;	PR 03-AUG-1999;	99US-0146388.
XX XX 14-MAY-1999;	PR 04-AUG-1999;	99US-0146389.
XX XX 14-MAY-1999;	PR 05-AUG-1999;	99US-0147038.
XX XX 14-MAY-1999;	PR 06-AUG-1999;	99US-0147204.
XX XX 14-MAY-1999;	PR 07-AUG-1999;	99US-0147302.
XX XX 14-MAY-1999;	PR 08-AUG-1999;	99US-0147192.
XX XX 14-MAY-1999;	PR 09-AUG-1999;	99US-0147260.
XX XX 14-MAY-1999;	PR 10-AUG-1999;	99US-0147303.
XX XX 14-MAY-1999;	PR 11-AUG-1999;	99US-0147416.
XX XX 14-MAY-1999;	PR 12-AUG-1999;	99US-0147935.
XX XX 14-MAY-1999;	PR 13-AUG-1999;	99US-0148171.
XX XX 14-MAY-1999;	PR 14-AUG-1999;	99US-0148319.
XX XX 14-MAY-1999;	PR 15-AUG-1999;	99US-0148341.
XX XX 14-MAY-1999;	PR 16-AUG-1999;	99US-0148565.
XX XX 14-MAY-1999;	PR 17-AUG-1999;	99US-0148664.
XX XX 14-MAY-1999;	PR 18-AUG-1999;	99US-0149175.
XX XX 14-MAY-1999;	PR 19-AUG-1999;	99US-0149426.
XX XX 14-MAY-1999;	PR 20-AUG-1999;	99US-0149722.
XX XX 14-MAY-1999;	PR 21-AUG-1999;	99US-0149723.
XX XX 14-MAY-1999;	PR 22-AUG-1999;	99US-0149929.
XX XX 14-MAY-1999;	PR 23-AUG-1999;	99US-0149902.
XX XX 14-MAY-1999;	PR 24-AUG-1999;	99US-0149930.
XX XX 14-MAY-1999;	PR 25-AUG-1999;	99US-0150566.
XX XX 14-MAY-1999;	PR 26-AUG-1999;	99US-0150884.
XX XX 14-MAY-1999;	PR 27-AUG-1999;	99US-0151065.
XX XX 14-MAY-1999;	PR 28-AUG-1999;	99US-0151066.
XX XX 14-MAY-1999;	PR 29-AUG-1999;	99US-0151080.
XX XX 14-MAY-1999;	PR 30-AUG-1999;	99US-0151303.
XX XX 14-MAY-1999;	PR 31-AUG-1999;	99US-0151438.
XX XX 14-MAY-1999;	PR 01-SEP-1999;	99US-0151930.
XX XX 14-MAY-1999;	PR 02-SEP-1999;	99US-0152363.
XX XX 14-MAY-1999;	PR 03-SEP-1999;	99US-0153070.
XX XX 14-MAY-1999;	PR 04-SEP-1999;	
XX XX 14-MAY-1999;	PR 05-SEP-1999;	
XX XX 14-MAY-1999;	PR 06-SEP-1999;	
XX XX 14-MAY-1999;	PR 07-SEP-1999;	
XX XX 14-MAY-1999;	PR 08-SEP-1999;	
XX XX 14-MAY-1999;	PR 09-SEP-1999;	
XX XX 14-MAY-1999;	PR 10-SEP-1999;	

```

PR 13-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      82.2%; Score 14.8; DB 21; Length 1554;
Best Local Similarity 88.9%; Pred. No. 2.8e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atactctgtcatgttaa 18
   || ||||| |||||
Db 264 atcctctgttatgttaa 281

RESULT 12
ABLT25967/c
ID ABL25967 standard; DNA; 1926 BP.
XX AC ABL25967;
XX AC
XX DT 26-MAR-2002 (first entry)
XX DE
XX KW Drosophila melanogaster genomic polynucleotide SEQ ID NO 29374.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ds.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX XX
XX PD 27-SEP-2001.
XX PF
XX 23-MAR-2001; 2001WO-US09231.

```

```

XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX (PEKE ) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
DR
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Claim 1; SEQ ID NO 29374; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA
CC sequences (ABLI01840-ABLI6175) and the encoded proteins
CC (ABBS7737-ABBS2072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1926 BP; 394 A; 531 C; 565 G; 436 T; 0 other;

Query Match      82.2%; Score 14.8; DB 23; Length 1926;
Best Local Similarity 88.9%; Pred. No. 2.8e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atactctgtcatgttaa 18
   || ||||| |||||
Db 1570 ACACGTGTGTCATGTAA 1553

RESULT 13
AAZ30413
ID AAZ30413 standard; DNA; 1935 BP.
XX AC
XX AC AAZ30413;
XX DT 21-DEC-1999 (first entry)
XX DE
XX DE Arabidopsis thaliana Mlo fungal resistance gene CIB10295.
XX KW
XX KW Consensus; resistance; fungus; pathogen; wheat; cell wall; apposition;
XX KW papilla; contact site; callose; carbohydrate; phenol; transgenic plant;
XX KW Mlo; Erysiphe graminis; powdery mildew; ss.
XX OS Arabidopsis thaliana.
XX XX
XX FH Key Location/Qualifiers
XX CDS 79..1810
XX FT /*tag= a
XX FT /product= "Mlo fungal resistance protein"
XX PN WO9947552-A2.
XX XX
XX PD 23-SEP-1999.
XX PF
XX PF 17-MAR-1999; 99WO-EP01779.
XX PF
XX PR 17-MAR-1998; 98US-0042763.
XX XX
XX (NOVS ) NOVARTIS AG.
XX PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX XX
XX PI Salmeron JM, Weislo LJ, Strawn LJ, Kramer CM;

```

XX WPI: 1999-571820/48.
 DR P-PSDB; AAY26970.
 XX
 PT New proteins useful for generating transgenic plants resistant to
 PT fungal infection -
 XX
 PS Claim 3; Page 78-81; 102pp; English.
 XX
 XX This sequence represents the coding region for the Arabidopsis thaliana
 CC fungal resistance gene CIB10295. The Mlo protein confers resistance to
 CC fungal pathogens by stimulating the formation of large cell wall
 CC appositions, designated papillae, at the contact site with the fungal
 CC pathogen. These papillae mainly contain callose, but also contain
 CC carbohydrate, phenols and proteins and are used to prevent penetration
 CC of the fungal hyphae into the plant. The new Mlo sequences are used to
 CC generate transgenic plants resistant to fungal pathogens, especially
 CC Erysiphe graminis (powdery mildew).
 CC
 XX Sequence 1935 BP; 576 A; 382 C; 398 G; 579 T; 0 other;
 SQ

Query Match 82.2%; Score 14.8; DB 20; Length 1935;
 Best Local Similarity 88.9%; Pred. No. 2.8e+02;
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atactctggctcatgtaa 18
 |||||
 DB 1630 atactctggctcatgtaa 1647

RESULT 14
 AAF24587
 ID AAF24587 standard; DNA; 1935 BP.
 AC
 XX
 XX AAF24587;
 DT 20-APR-2001 (first entry)
 XX
 XX Nucleotide sequence of the Arabidopsis Mlo protein CIB10295.
 DE
 XX Wheat; fungal resistance; fungal pathogen; Erysiphe graminis; papillae;
 KW powdery mildew; Mlo; cell wall apposition; transgenic plant; ss.
 XX Arabidopsis thaliana.
 OS
 XX Key Location/Qualifiers
 FH 89..1810
 ET CDS /*tag= a
 FT /product= "Mlo protein"
 FT
 XX W0200078799-A2.
 XX
 XX 28-DEC-2000.
 XX
 XX 16-JUN-2000; 2000WO-EP05576.
 PF
 XX
 XX 18-JUN-1999; 99US-0336112.
 XX
 XX (NOVS) NOVARTIS AG.
 PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GFS MBH.
 XX
 XX Salmeron JM, Weislo LJ, Strawn LJ, Kramer CM, Wang HX;
 PI Vernoolj BT, Levin JZ, Heifetz PB, Patton DA, Que Q;
 XX
 XX WPI: 2001-112311/12.
 DR P-PSDB; AAB31234.
 XX
 XX Novel polynucleotide which encodes Mlo protein from wheat, useful for
 PT producing fungal resistant plants, in particular wheat plant -
 XX
 XX Claim 33; Page 106-108; 128pp; English.
 PS
 XX

CC The present sequence encodes a Mlo protein. Mlo proteins given
 CC resistance to fungal pathogens which infect living epidermal plant cells.
 CC Mlo proteins give proteins resistance to fungal pathogens, especially
 CC Erysiphe graminis (powdery mildews). In barley, mutations at the Mlo
 CC locus are responsible for a plant resistant to fungal pathogens. The mechanism
 CC of Mlo resistance involves the formation of large cell wall appositions,
 CC called papillae, at pathogen contact sites. Papillae mainly contain
 CC callose, but also carbohydrates, phenols and proteins. Polynucleotides
 CC encoding Mlo proteins are useful for making transgenic plants,
 CC preferably wheat, which are resistant to fungal pathogens.
 XX
 XX Sequence 1935 BP; 576 A; 382 C; 398 G; 579 T; 0 other;
 SQ

Query Match 82.2%; Score 14.8; DB 22; Length 1935;
 Best Local Similarity 88.9%; Pred. No. 2.8e+02;
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atactctggctcatgtaa 18
 |||||
 DB 1630 atactctggctcatgtaa 1647

RESULT 15
 AAK67740
 ID AAK67740 standard; DNA; 1943 BP.
 AC
 XX
 XX AAK67740;
 DT 06-NOV-2001 (first entry)
 XX
 XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:22552.
 DE
 XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
 KW cytostatic; gene therapy; vaccine; metastasis; ds.
 XX
 XX Homo sapiens.
 OS
 XX W0200157182-A2.
 PN
 XX
 XX 09-AUG-2001.
 PD
 XX
 XX 17-JAN-2001; 2001WO-US01354.
 PF
 XX
 XX 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR

PR	14-AUG-2000;	2000US-02255759;
PR	18-AUG-2000;	2000US-02262679;
PR	22-AUG-2000;	2000US-02266811;
PR	22-AUG-2000;	2000US-02266868;
PR	22-AUG-2000;	2000US-0227102;
PR	23-AUG-2000;	2000US-0227009;
PR	30-AUG-2000;	2000US-0228924;
PR	01-SEP-2000;	2000US-02292837;
PR	01-SEP-2000;	2000US-02293443;
PR	01-SEP-2000;	2000US-02293445;
PR	05-SEP-2000;	2000US-02293509;
PR	06-SEP-2000;	2000US-0229513;
PR	06-SEP-2000;	2000US-0230438;
PR	08-SEP-2000;	2000US-02311243;
PR	08-SEP-2000;	2000US-02311242;
PR	08-SEP-2000;	2000US-02311413;
PR	08-SEP-2000;	2000US-0231414;
PR	08-SEP-2000;	2000US-0232080;
PR	08-SEP-2000;	2000US-0232081;
PR	12-SEP-2000;	2000US-0231968;
PR	14-SEP-2000;	2000US-0232397;
PR	14-SEP-2000;	2000US-0232398;
PR	14-SEP-2000;	2000US-0232399;
PR	14-SEP-2000;	2000US-0232400;
PR	14-SEP-2000;	2000US-0232401;
PR	14-SEP-2000;	2000US-0232403;
PR	14-SEP-2000;	2000US-0233063;
PR	14-SEP-2000;	2000US-0233064;
PR	14-SEP-2000;	2000US-0233065;
PR	21-SEP-2000;	2000US-0234223;
PR	21-SEP-2000;	2000US-0234274;
PR	25-SEP-2000;	2000US-0234997;
PR	25-SEP-2000;	2000US-0234998;
PR	26-SEP-2000;	2000US-0235484;
PR	27-SEP-2000;	2000US-0235834;
PR	29-SEP-2000;	2000US-0235836;
PR	29-SEP-2000;	2000US-0236327;
PR	29-SEP-2000;	2000US-0236367;
PR	29-SEP-2000;	2000US-0236368;
PR	29-SEP-2000;	2000US-0236369;
PR	29-SEP-2000;	2000US-0236370;
PR	02-OCT-2000;	2000US-0236802;
PR	02-OCT-2000;	2000US-0237037;
PR	02-OCT-2000;	2000US-0237038;
PR	02-OCT-2000;	2000US-0237039;
PR	13-OCT-2000;	2000US-0237040;
PR	13-OCT-2000;	2000US-0239935;
PR	13-OCT-2000;	2000US-0239937;
PR	20-OCT-2000;	2000US-0240960;
PR	20-OCT-2000;	2000US-0241221;
PR	20-OCT-2000;	2000US-0241785;
PR	20-OCT-2000;	2000US-0241786;
PR	20-OCT-2000;	2000US-0241787;
PR	20-OCT-2000;	2000US-0241808;
PR	20-OCT-2000;	2000US-0241809;
PR	01-NOV-2000;	2000US-0241826;
PR	08-NOV-2000;	2000US-0246474;
PR	08-NOV-2000;	2000US-0246475;
PR	08-NOV-2000;	2000US-0246476;
PR	08-NOV-2000;	2000US-0246477;
PR	08-NOV-2000;	2000US-0246478;
PR	08-NOV-2000;	2000US-0246523;
PR	08-NOV-2000;	2000US-0246524;
PR	08-NOV-2000;	2000US-0246525;
PR	08-NOV-2000;	2000US-0246526;
PR	08-NOV-2000;	2000US-0246527;
PR	08-NOV-2000;	2000US-0246528;
PR	08-NOV-2000;	2000US-0246532;
PR	08-NOV-2000;	2000US-0246601;
PR	08-NOV-2000;	2000US-0246610;

PR	08-NOV-2000;	2000US-02466111;
PR	08-NOV-2000;	2000US-02466133;
PR	17-NOV-2000;	2000US-02462307;
PR	17-NOV-2000;	2000US-02492008;
PR	17-NOV-2000;	2000US-02492008;
PR	17-NOV-2000;	2000US-02492010;
PR	17-NOV-2000;	2000US-02492112;
PR	17-NOV-2000;	2000US-02492112;
PR	17-NOV-2000;	2000US-02492114;
PR	17-NOV-2000;	2000US-02492115;
PR	17-NOV-2000;	2000US-02492116;
PR	17-NOV-2000;	2000US-02492117;
PR	17-NOV-2000;	2000US-02492118;
PR	17-NOV-2000;	2000US-02492124;
PR	17-NOV-2000;	2000US-02492455;
PR	17-NOV-2000;	2000US-02492456;
PR	17-NOV-2000;	2000US-02492455;
PR	17-NOV-2000;	2000US-02492597;
PR	17-NOV-2000;	2000US-02492599;
PR	17-NOV-2000;	2000US-02493000;
PR	01-DEC-2000;	2000US-02493000;
PR	01-DEC-2000;	2000US-02503391;
PR	05-DEC-2000;	2000US-02510300;
PR	05-DEC-2000;	2000US-02519888;
PR	05-DEC-2000;	2000US-02567179;
PR	06-DEC-2000;	2000US-02514779;
PR	08-DEC-2000;	2000US-02518586;
PR	08-DEC-2000;	2000US-02518668;
PR	08-DEC-2000;	2000US-02518669;
PR	08-DEC-2000;	2000US-02519889;
PR	11-DEC-2000;	2000US-02519900;
PR	11-DEC-2000;	2000US-02540097;
PR	05-JAN-2001;	2001US-02549678;
X	(HUMA-)	HUMAN GENOME SCI INC.

Query Match	82.2%;	Score 14.8;	DB 22;	Length 1943;
Best Local Similarity	88.9%;			
			Pred. No. 2.8e+02;	

Qy 1 atactctgggtcatgttaa 18

us-09-899-718a-8.rng

Thu Aug 1 08:30:07 2002

Db 474 atacttggtcatgttaa 491
||||| || |||||||||

Search completed: July 31, 2002, 14:11:28
Job time: 17385 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 14:02:11 ; Search time 6034.22 Seconds
(without alignments)
62.424 Million cell updates/sec

Title: US-09-899-718A-8

Perfect score: 18

Sequence: 1 atactctggctcatgttaa 18

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 1000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sv.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pi.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
-----	-----	-----	-----	-----	-----

1	18	100.0	18	6	AX349070
2	18	100.0	3785	5	AX349063
3	16.4	91.1	51019	5	AL591420
4	16.4	91.1	55613	9	AL591420
5	16.4	91.1	73309	2	AL606722
6	16.4	91.1	81017	2	AL606722
7	16.4	91.1	105939	2	AC025010
8	16.4	91.1	110000	2	AC094963
9	16.4	91.1	137140	2	AL359456-1
10	16.4	91.1	138716	2	AC091696
11	16.4	91.1	147267	2	AL591028
12	16.4	91.1	148418	9	AC097356
13	16.4	91.1	164235	9	AC020698
14	16.4	91.1	168922	30	AC036175
15	16.4	91.1	176355	2	AC108072
16	16.4	91.1	182848	9	AL356053
17	16.4	91.1	184672	2	AC106149
18	16.4	91.1	186848	2	AP002508
19	16.4	91.1	190160	2	AC009272
20	16.4	91.1	190376	2	AF001837
21	16.4	91.1	197069	2	AC090229
22	16.4	91.1	198540	9	AC090370
23	16.4	91.1	229041	2	AC099697
24	16	88.9	19532	9	AL596243
25	16	88.9	92496	9	AL589848
26	16	88.9	126247	2	AC091955
27	16	88.9	148295	2	AL662913
28	16	88.9	149988	2	AC011929
29	16	88.9	154618	2	AC073403
30	16	88.9	164180	9	AC016725
31	16	88.9	171692	2	AL662896
32	16	88.9	174472	2	AC094034
33	16	88.9	196848	9	AC007684
34	15.4	85.6	1709	10	AF176522
35	15.4	85.6	1770	5	CCZP2A
36	15.4	85.6	1897	5	CCZP2C
37	15.4	85.6	1873	10	AB041571
38	15.4	85.6	2826	4	BTA250379
39	15.4	85.6	3295	5	CCZP2GEN
40	15.4	85.6	41947	9	HSAC000358
41	15.4	85.6	43284	9	AC004201
42	15.4	85.6	50000	8	AY013245
43	15.4	85.6	53949	2	AC090724
44	15.4	85.6	65020	2	AC102336
45	15.4	85.6	67895	2	AP004160

ALIGNMENTS

RESULT	1	AX349070	Sequence 8 from Patent WO0202785.	18 bp	DNA	linear	PAT 06-FEB-2002
LOCUS	AX349070	Sequence 8 from Patent WO0202785.					
DEFINITION	AX349070	Sequence 8 from Patent WO0202785.					
ACCESSION	AX349070	Sequence 8 from Patent WO0202785.					
VERSION	AX349070.1	GI:18615105					
KEYWORDS		synthetic construct.					
SOURCE		synthetic construct.					
ORGANISM		synthetic construct.					
REFERENCE		1 (sites)					
AUTHORS		Sprunck S., Kluth A., Becker D., Luetticke S. and Loerz H.					
TITLE		Promoters of gene expression in plant caryopses					
JOURNAL		Patent: WO 0202785-A 8 10-JAN-2002;					
FEATURES		Aventis CropScience GmbH (DE)					
source		Location/Qualifiers					
		1. .18					
		/organism="synthetic construct"					
		/db_xref="taxon:32630"					
		/note="Oligonucleotide"					
		5 a 3 c 3 g 7 t					

Thu Aug 1 08:30:06 2002

```

Query Match      100.0%; Score 18; DB 6; Length 18;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atactctgggtcatgttaa 18
    |||||
Db 1 ATACTCTGGTCATGTAA 18

RESULT 2
AX349063 AX349063 3785 bp DNA linear PAT 06-FEB-2002
LOCUS
DEFINITION Sequence 1 from Patent WO0202785.
ACCESSION AX349063
VERSION AX349063.1 GI:18615098
KEYWORDS bread wheat.
SOURCE
ORGANISM
Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
REFERENCE
1 (sites) Sprunck,S., Kluth,A., Becker,D., Luetticke,S. and Loerz,H.
PROMOTERS OF GENE EXPRESSION IN PLANT CARYOPSES
TITLE Patent: WO 0202785-A 1 10-JAN-2002;
JOURNAL Aventis CropScience GmbH (DE)
FEATURES
Location/Qualifiers
Source
1..3785
/organism="Triticum aestivum"
/db_xref="taxon:4565"
BASE COUNT 945 a 980 c 899 g 961 t

Query Match      100.0%; Score 18; DB 6; Length 3785;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atactctgggtcatgttaa 18
    |||||
Db 2999 ATACTCTGGTCATGTAA 3016

RESULT 3
AL591420 51019 bp DNA linear VRT 23-JAN-2002
LOCUS
DEFINITION Zebrafish DNA sequence from clone XX-221H6 on chromosome 7 Contains
a novel gene similar to nitr3r.1 (novel immune-type receptor 3r.1),
a novel gene similar to nitr3 and part of two novel genes similar
to nitr's, complete sequence.
ACCESSION AL591420
VERSION AL591420.5 GI:16973933
KEYWORDS HFG; immune-type receptor; nitr3; nitr3r.1.
SOURCE zebrafish.
ORGANISM
Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 51019)
Hammond,S.
Direct Submission
Submitted (10-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humayun@sanger.ac.uk
On Nov 17, 2001 this sequence version replaced gi:15722150.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all

```

```

regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em., EMBL; Sw.,
SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep
a Zebrafish PAC library
VECTOR: pCYPAC-6
IMPORTANT: This sequence is not the entire insert of clone XX-221H6
It may be shorter because we sequence overlapping sections only
once, except for a short overlap.
The true right end of clone XX-221H6 is at 51019 in this sequence.
The true right end of clone XX-139F19 is at 2000 in this sequence.
This clone was isolated from the BUSM1 library (C. Amemiya) and
provided by C. Amemiya (Amemiya lab, Virginia Mason Research
Center, Seattle, USA).
Location/Qualifiers
Source
1..51019
/organism="Danio rerio"
/db_xref="taxon:7955"
/chromosome="7"
/clone="XX-221H6"
/clone_lib="Amemiya"
repeat_region 567..630
/note="16 copies 4 mer at 75% conserved"
repeat_region 588..631
/note="22 copies 2 mer ta 86% conserved"
repeat_region 1156..1263
/note="9 copies 12 mer 86% conserved"
repeat_region 1157..1256
/note="50 copies 2 mer ta 91% conserved"
repeat_region 1163..1260
/note="7 copies 14 mer 89% conserved"
repeat_region 1165..1256
/note="23 copies 4 mer tata 91% conserved"
gene join(2780..2825,3152..3505,4439..4573,5294..5328,
5674..5796)
CDS join(2780..2825,3152..3505,4439..4573,5294..5328,
5674..5796)
/gene="SC:d221H6.1"
/note="match: cDNAs: Em:AF318401
Tr:CAD12598 Tr:AAK60113"
match: proteins: Tr:CAD12598 Tr:AAK60113"
/evidence=not_experimental
/product="SC:d221H6.1 (novel immune-type receptor similar
to nitr3r.1)"
/protein_id="CAD21620.1"
/db_xref="GI:18369621"
/translation="MSLQDCEFFLLTAYGTCEDFIHQPLVVAELGSSVTLPCPH
SDDFITTSWYKHSACKPLLIASPNQSVTVQVAENNTNRFITTAGSNLSIL
HLEDEFPANYCAKFLNLMFPGETILLNEDRNISTSVSPSFPTVPVVCILLIIS
VTSVMNVLIVIQKSRKKTETQLRSQINQIKLDDLYAALHFSKTKPTTSRSSMKT
IOETIYSETVH"
repeat_region 3041..3144
/note="52 copies 2 mer tt 62% conserved"
misc_feature 4000..4019
/gene="SC:d221H6.1"
/note="Inconsistency in the number of bases in
mononucleotide run between subclones."
4019..4060
/note="3 copies 14 mer 100% conserved"
repeat_region complement(4020..4061)
/misc_feature /note="random repeat. Inconsistency in the number of
copies of the repeat element between subclones."
5902..5959
repeat_region /note="29 copies 2 mer at 81% conserved"
7894..7919

```



```

repeat_region /note="13 copies 2 mer tt 100% conserved"
8572..8637
gene /note="33 copies 2 mer tt 68% conserved"
8641..8994
CDS /gene="SC:d2221H6.2"
<8641..>8994
/note="SC:d2221H6.2"
match: proteins: Tr:CAD12524 Tr:CAD12506 Tr:CAD12582
Tr:CAD12581"
/codon_start=3
/evidence=not_experimental
/product="SC:d2221H6.2 (novel immune-type receptor)"
/protein_id="CAD21621.1"
/db_xref="GI:18369622"
/translation="GTCEEDFIHQOPLVLAELGDSVTLPCHSDFVTYISWKHSAG
KPLLIAYSDNSGVTYRNFAFNNTFFITVAGSYNLTIIHLEKEDFATYVCVKDF
LNLFMGEGTILLCK"
9391..9371
/note="DANA repeat: matches 322..405 of consensus"
Join(10373..10727,10812..10913)
/gene="SC:d2221H6.3"
Join(<10373..10727,10812..>10913)
/gene="SC:d2221H6.3"
/note="predicted by Pfam
match: proteins: Tr:AAK60111 Tr:AAK60112 Tr:CAD12599
Tr:CAD12515"
/codon_start=1
/evidence=not_experimental
/product="SC:d221H6.3 (novel immune-type receptor)"
/protein_id="CAD21622.1"
/db_xref="GI:18369623"
/translation="ETDRISSTVIOQPSDRLHPGDSVTLQCSVSHTCAGHYRVY
FKHSGSQPIIYTHNRSQDLESSEKSFVQSCVYLSQTELTTSDAGVYCAVD
TCGRFGNGTKLIERSSLPWVLLIFVISAVIIVFLIIRYCKE"
13097..13274
/note="match: STS: Em:G40916"
14549..14869
/note="DANA repeat: matches 10..377 of consensus"
16279..16574
/note="DANA repeat: matches 89..406 of consensus"
16588..16622
/note="DANA repeat: matches 5..59 of consensus"
17443..17464
/note="11 copies 2 mer tt 100% conserved"
Join(17884..17929,18244..18597,20317..20670,20748..20846,
21489..21523,21967..22089)
/gene="SC:d2221H6.4"
Join(17884..17929,18244..18597,20317..20670,20748..20846,
21489..21523,21967..22089)
/note="match: CDNAS: Em:AF318398 Em:AF318400 Em:AF318399
match: proteins: Tr:AAK60106 Tr:CAD12509 Tr:CAD12585
Tr:AAK60111"
Tr:AAK60112 Tr:CAD12599
/codon_start=1
/evidence=not_experimental
/product="SC:d221H6.4 (novel immune-type receptor similar
to nitr3)"
/db_xref="GI:18369624"
/translation="MSLQSSFTFFELIFAYGTCERDFTHQOPLVLAELGSRVTLPCPH
SDIVYTVSWTKHSAKKPLLIAYSDNSGVTYONAFNNTNRFITVAGSYNLTII
HLEKEDFATYVCVKDFLNLFMGEGTILLRKETDRISSTVIOQPSDRLHPGDSVTL
QCSVSHICAGHYRVYFKRSGYSQPGFIYTHNRSQDLESSEKSFVQSCVYLS
QTELTTSDAGVYCAVDTCGKTFHCNGTKLIEASSFWNPVVVILFTISVTIVANIV
LVLONRRKETTSLQGLNQIETDLDLTAALHFSTKPTSPSTRSSIKTIQETIYSETT
VH"
18177..18240
/note="32 copies 2 mer tt 70% conserved"
19173..19252
/note="DANA repeat: matches 322..405 of consensus"
19325..19382

```

```

repeat_region /note="29 copies 2 mer ga 69% conserved"
19721..19776
repeat_region /note="4 copies 14 mer 80% conserved"
21020..21227
/note="TDR1 repeat: matches 727..934 of consensus"
21332..21377
repeat_region /note="23 copies 2 mer aa 80% conserved"
21756..21803
/note="4 copies 12 mer 91% conserved"
22349..22528
/note="match: STS: Em:G39804"
22354..22387
/note="17 copies 2 mer ca 97% conserved"
22702..22867
/note="match: STS: Em:G41123"
22913..22957
/note="23 copies 2 mer ta 80% conserved"
23406..23531
/note="6 copies 21 mer 70% conserved"
23996..24067
/note="6 copies 12 mer 100% conserved"
25897..25968
/note="36 copies 2 mer aa 66% conserved"
27596..27827
/note="DANA repeat: matches 147..406 of consensus"
27825..27883
/note="DANA repeat: matches 1..59 of consensus"

Query Match 91.1%; Score 16.4; DB 5; Length 51019;
Best Local Similarity 94.4%; Pred. No. 90;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atactctggtcatgtaa 18
|||||
Db 42115 ATACTCTGTAATGTAA 42132

RESULT 4
AL512443/c
LOCUS Human DNA sequence from clone RP11-108P2 on chromosome 1, complete
DEFINITION sequence.
AL512443
ACCESSION AL512443.7 GI:16508240
VERSION HTG.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 55613)
AUTHORS Frankland,J
TITLE Direct Submission
JOURNAL Submitted (26-Oct-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Oct 29, 2001 this sequence version replaced gi:14529902.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em:, EMBL; Sw:,
SWISSPROT; Tr:, TrEMBL; Wp:, WormPEP; information on the WormPEP
database can be found at

```

Thu Aug 1 08:30:06 2002

us-09-899-718a-8.rge

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1>
 RP11-108P2 is from the library RPCR-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
 VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-108P2 it may be shorter because we sequence overlapping sections only once, except for a short overlap.
 The true left end of clone RP11-38609 is at 53614 in this sequence.
 The true right end of clone RP3-333A15 is at 2000 in this sequence.

FEATURES

source
 Location/Qualifiers
 1..55613
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-108P2"
 /clone_lib="RPCI-11.1"
 18193 a 10125 c 9972 g 17323 t

BASE COUNT

Query Match 91.1%; Score 16.4; DB 9; Length 55613;
 Best Local Similarity 94.4%; Pred. No. 91;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 atactctgcatgttaa 18

||||| |||||||||

Db 31634 ATACTCTGTCATGTAA 31617

RESULT

AL606722/c 5
 LOCUS 73309 bp DNA linear HTG 20-DEC-2001
 DEFINITION Danio rerio chromosome 1 clone XX-27N24, *** SEQUENCING IN PROGRESS
 ***, in unordered pieces.

ACCESSION AL606722
 VERSION AL606722.2 GI:17976610
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE zebrafish

ORGANISM

Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.

1 (sites)
 Burton, J.

Direct Submission
 Submitted (19-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Dec 21, 2001 this sequence version replaced gi:15626247.

COMMENT

----- Genome Center
 Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: d227N24

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 72055 bases at least Q40

Consensus quality: 72273 bases at least Q30

Consensus quality: 72397 bases at least Q20

Insert size: 72609; sum-of-contigs

Insert size: 88868; 5.6% error; agarose-fp

Quality coverage: 11.34x in Q20 bases; sum-of-contigs Quality

coverage: 9.50x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

FEATURES

source
 Location/Qualifiers
 1..73309
 /organism="Danio rerio"
 /db_xref="taxon:7955"
 /chromosome="1"
 /clone="XX-27N24"
 /clone_lib="Amemiya"
 1..3149
 /note="assembly fragment:01316
 fragment_chain:1"
 3250..14161
 /note="assembly fragment:00988
 fragment_chain:1"
 14262..25306
 /note="assembly fragment:01324
 fragment_chain:1"
 25407..29595
 /note="assembly fragment:01886
 fragment_chain:1"
 29696..32697
 /note="assembly fragment:01272
 fragment_chain:1"
 32798..52684
 /note="assembly fragment:01855
 fragment_chain:1"
 52785..66125
 /note="assembly fragment:01033
 fragment_chain:1"
 66226..73309
 /note="assembly fragment:00736"
 21551 a 14405 c 13666 g 22985 t 702 others

ORIGIN

Query Match 91.1%; Score 16.4; DB 2; Length 73309;
 Best Local Similarity 94.4%; Pred. No. 92;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 atactctgcatgttaa 18

||||| |||||||||

Db 32283 ATACTCTGTCATGTAA 32266

RESULT

AC025010 6
 LOCUS 81017 bp DNA linear HTG 13-JUL-2000
 DEFINITION Homo sapiens chromosome 4 clone RP11-85D17 map 4, LOW-PASS SEQUENCE
 SAMPLING.

ACCESSION AC025010
 VERSION AC025010.1 GI:7145054

KEYWORDS HTG; HTGS_PHASE0.

SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 81017)

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens chromosome 4, clone RP11-85D17

Unpublished

2 (bases 1 to 81017)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,

Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,

Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,

Collamore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S.,

Dodge, S., Domino, M., Doyle, M., Ferrelira, P., Fitzhugh, W., Gage, D.,

Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,

Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,

Howland, J.C., Illiev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,

Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J.,
 Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
 Meldrum, J., Meneus, L., Minova, I., Miranda, C., Mienga, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
 Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE
 JOURNAL
 COMMENT

Submitted (03-MAR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Center project name: L7756

Center clone name: 85_D_17

* NOTE: This record contains 94 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1 709: contig of 709 bp in length
 710 809: gap of 100 bp
 810 1567: contig of 758 bp in length
 1568 1667: gap of 100 bp
 1668 2433: contig of 766 bp in length
 2434 2533: gap of 100 bp
 2534 3310: contig of 777 bp in length
 3311 3410: gap of 100 bp
 3411 4176: contig of 766 bp in length
 4177 4276: gap of 100 bp
 4277 5036: contig of 760 bp in length
 5037 5136: gap of 100 bp
 5137 5899: contig of 763 bp in length
 5900 5999: gap of 100 bp
 6000 6764: contig of 765 bp in length
 6765 6864: gap of 100 bp
 6865 7612: contig of 748 bp in length
 7613 7712: gap of 100 bp
 7713 8467: contig of 755 bp in length
 8468 8567: gap of 100 bp
 8568 9320: contig of 753 bp in length
 9321 9420: gap of 100 bp
 9421 10187: contig of 767 bp in length
 10188 10287: gap of 100 bp
 10288 11042: contig of 755 bp in length
 11043 11142: gap of 100 bp
 11143 11910: contig of 768 bp in length
 11911 12010: gap of 100 bp
 12011 12757: contig of 747 bp in length
 12758 12857: gap of 100 bp
 12858 13619: contig of 762 bp in length
 13620 13719: gap of 100 bp
 13720 14493: contig of 774 bp in length
 14494 14593: gap of 100 bp
 14594 15348: contig of 755 bp in length

15349 15448: gap of 100 bp
 15449 16206: contig of 758 bp in length
 16207 16306: gap of 100 bp
 16307 17054: contig of 748 bp in length
 17055 17154: gap of 100 bp
 17155 17900: contig of 746 bp in length
 17901 18000: gap of 100 bp
 18001 18767: contig of 767 bp in length
 18768 18867: gap of 100 bp
 18868 19642: contig of 775 bp in length
 19643 19742: gap of 100 bp
 19743 20493: contig of 751 bp in length
 20494 20593: gap of 100 bp
 20594 21337: contig of 744 bp in length
 21338 21437: gap of 100 bp
 21438 22202: contig of 765 bp in length
 22203 22302: gap of 100 bp
 22303 23072: contig of 770 bp in length
 23073 23172: gap of 100 bp
 23173 23941: contig of 769 bp in length
 23942 24041: gap of 100 bp
 24042 24799: contig of 758 bp in length
 24800 24899: gap of 100 bp
 24900 25667: contig of 768 bp in length
 25668 25767: gap of 100 bp
 25768 26515: contig of 748 bp in length
 26516 26615: gap of 100 bp
 26616 27375: contig of 760 bp in length
 27376 27475: gap of 100 bp
 27476 28238: contig of 763 bp in length
 28239 28338: gap of 100 bp
 28339 29104: contig of 766 bp in length
 29105 29204: gap of 100 bp
 29205 29969: contig of 765 bp in length
 29970 30069: gap of 100 bp
 30070 30837: contig of 768 bp in length
 30838 30937: gap of 100 bp
 30938 31710: contig of 773 bp in length
 31711 31810: gap of 100 bp
 31811 32576: contig of 766 bp in length
 32577 32676: gap of 100 bp
 32677 33461: contig of 785 bp in length
 33462 33561: gap of 100 bp
 33562 34337: contig of 776 bp in length
 34338 34437: gap of 100 bp
 34438 35216: contig of 779 bp in length
 35217 35316: gap of 100 bp
 35317 36082: contig of 766 bp in length
 36083 36182: gap of 100 bp
 36183 36951: contig of 769 bp in length
 36952 37051: gap of 100 bp
 37052 37815: contig of 764 bp in length
 37816 37915: gap of 100 bp
 37916 38676: contig of 761 bp in length
 38677 38776: gap of 100 bp
 38777 39563: contig of 787 bp in length
 39564 39663: gap of 100 bp
 39664 40445: contig of 782 bp in length
 40446 40545: gap of 100 bp
 40546 41316: contig of 771 bp in length
 41317 41416: gap of 100 bp
 41417 42181: contig of 765 bp in length
 42182 42281: gap of 100 bp
 42282 43030: contig of 749 bp in length
 43031 43130: gap of 100 bp
 43131 43913: contig of 783 bp in length
 43914 44013: gap of 100 bp
 44014 44770: contig of 757 bp in length
 44771 44870: gap of 100 bp
 44871 45631: contig of 761 bp in length
 45632 45731: gap of 100 bp
 45732 46487: contig of 756 bp in length
 46488 46587: gap of 100 bp

Thu Aug 1 08:30:06 2002

* 46588 47340: contig of 753 bp in length
 * 47341 47440: gap of 100 bp
 * 47441 48171: contig of 731 bp in length
 * 48172 48271: gap of 100 bp
 * 48272 49028: contig of 757 bp in length
 * 49029 49128: gap of 100 bp
 * 49129 49883: contig of 755 bp in length
 * 49884 49983: gap of 100 bp
 * 50748 50748: contig of 765 bp in length
 * 50749 50848: gap of 100 bp
 * 50849 51619: contig of 771 bp in length
 * 51620 51719: gap of 100 bp
 * 51720 52473: contig of 754 bp in length
 * 52474 52573: gap of 100 bp
 * 52574 53327: contig of 754 bp in length
 * 53328 53427: gap of 100 bp
 * 53428 54198: contig of 771 bp in length
 * 54199 54298: gap of 100 bp
 * 54299 55054: contig of 756 bp in length
 * 55055 55154: gap of 100 bp
 * 55155 55921: contig of 767 bp in length
 * 55922 56021: gap of 100 bp
 * 56022 56781: contig of 760 bp in length
 * 56782 56881: gap of 100 bp
 * 56882 57638: contig of 757 bp in length
 * 57639 57738: gap of 100 bp
 * 57739 58492: contig of 754 bp in length
 * 58493 58592: gap of 100 bp
 * 58593 59364: contig of 772 bp in length
 * 59365 59464: gap of 100 bp

Query Match 91.1%; Score 16.4; DB 2; Length 81017;

Best Local Similarity 94.4%; Pred. NO. 92;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 atactctggctcattgttaa 18

Db 12578 AGACTCGTCATGTTAA 12595

RESULT 7
 AC094963/c
 LOCUS
 DEFINITION Rattus norvegicus clone CH230-6L20, *** SEQUENCING IN PROGRESS ***,
 56 unordered pieces.
 AC094963
 AC094963.3 GI:17941764
 HTG: HTGS PHASE1.
 Norway rat.

ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

1 (bases 1 to 105939)
 Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,
 Alsbrooks, S.L., Amaratunga, H.C., Are, J.R., Banks, T., Barbaria, J.,
 Benton, J., Blincoe, K., Blankenburg, K., Bonnin, D., Bouck, J.,
 Bowie, S., Brieve, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,
 Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, F.F.,
 Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
 Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
 Coyle, M.D., Dathorne, S.K., David, R., Davila, M.L., Davis, C.,
 Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,
 Denn, A.L., Ding, J., Dinh, H.H., Douthwaite, K.J., Draper, H.,
 Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,
 Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
 Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
 Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,
 Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,
 Hernandez, J., Hernandez, O., Hodgson, A., Hoques, M., Holloway, C.,
 Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,
 Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
 Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,

Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
 Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
 Loulsegad, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
 Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
 Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M.,
 Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
 Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
 Nguyen, A., Nguyen, N., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,
 Ogih, M., Okwuonu, G., Oragunye, N., Pickens, R., Primus, E., Pu, L.,
 Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.,
 Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojokan, I., Rolfe, M.,
 Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoshkari, N.,
 Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H.,
 Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,
 Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,
 Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,
 Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
 Watlington, S., Williams, G., Williamson, A., Wleczka, R., Wooden, S.,
 Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.

TITLE

JOURNAL

REFERENCE

AUTHORS

JOURNAL

COMMENT

Unpublished

2 (bases 1 to 105939)

Worley, K.C.

Direct Submission

Submitted (15-SEP-2001) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Dec 20, 2001 this sequence version replaced gi:17062175.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GBUV

Center clone name: CH230-6L20

----- Summary Statistics

Assembly program: Phrap; version 0.990329First call to

findPhrapList

Consensus quality: 82923 bases at least Q40

Consensus quality: 92369 bases at least Q30

Consensus quality: 99005 bases at least Q20

Estimated insert size: 70929; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-fp estimation

Quality coverage: 0.8x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 56 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 4041: contig of 4041 bp in length

* 4042: gap of unknown length

* 4142: contig of 4544 bp in length

* 8685: gap of unknown length

* 8686: contig of 3322 bp in length

* 12107: gap of unknown length

* 12108: contig of 3521 bp in length

* 15728: gap of unknown length

* 15729: contig of 4293 bp in length

* 15829: contig of 2605 bp in length

* 20122: gap of unknown length

* 20122: contig of 2605 bp in length

* 22827: gap of unknown length

* 22927: contig of 2145 bp in length

* 25071: gap of unknown length

* 25172: contig of 3486 bp in length

* 28657: gap of unknown length

* 28658: gap of unknown length

* 28758 31090: contig of 2333 bp in length
* 31091 gap of unknown length
* 31191 3330: contig of 2140 bp in length
* 3331 3330: gap of unknown length
* 3331 3330: contig of 1668 bp in length
* 3331 35098: gap of unknown length
* 35199 37556: contig of 2358 bp in length
* 37557 37656: gap of unknown length
* 37657 38712: contig of 1056 bp in length
* 38713 38812: gap of unknown length
* 38813 40216: contig of 1404 bp in length
* 40217 40316: gap of unknown length
* 40317 42189: contig of 1873 bp in length
* 42190 42289: gap of unknown length
* 42290 44826: contig of 2537 bp in length
* 44827 44926: gap of unknown length
* 44927 46654: contig of 1728 bp in length
* 46655 46754: gap of unknown length
* 46755 48630: contig of 1876 bp in length
* 48631 48730: gap of unknown length
* 48731 50244: contig of 1514 bp in length
* 50245 50344: gap of unknown length
* 50345 52054: contig of 1710 bp in length
* 52055 52154: gap of unknown length
* 52155 53878: contig of 1724 bp in length
* 53879 53978: gap of unknown length
* 53979 55533: contig of 1555 bp in length
* 55534 55633: gap of unknown length
* 55634 57398: contig of 1765 bp in length
* 57399 57498: gap of unknown length
* 57499 59018: contig of 1520 bp in length
* 59019 59118: gap of unknown length
* 59119 60951: contig of 1833 bp in length
* 60952 61051: gap of unknown length
* 61052 63791: contig of 2740 bp in length
* 63792 64936: contig of 1045 bp in length
* 64937 65036: gap of unknown length
* 65037 65333: contig of 1497 bp in length
* 65334 66333: gap of unknown length
* 66334 68166: contig of 1533 bp in length
* 68167 68266: gap of unknown length
* 68267 69395: contig of 1129 bp in length
* 69396 69495: gap of unknown length
* 69496 70721: contig of 1226 bp in length
* 70722 70821: gap of unknown length
* 70822 72079: contig of 1258 bp in length
* 72080 72179: gap of unknown length
* 72180 73301: contig of 1121 bp in length
* 73301 73400: gap of unknown length
* 73401 74409: contig of 1009 bp in length
* 74410 74509: gap of unknown length
* 74510 76335: contig of 1826 bp in length
* 76336 76435: gap of unknown length
* 76436 77982: contig of 1547 bp in length
* 77983 78082: gap of unknown length
* 78083 79545: contig of 1463 bp in length
* 79546 79645: gap of unknown length
* 79646 81116: contig of 1471 bp in length
* 81117 82544: contig of 1328 bp in length
* 82545 82644: gap of unknown length
* 82645 83693: contig of 1049 bp in length
* 83694 83793: gap of unknown length
* 83794 85175: contig of 1382 bp in length
* 85176 85275: gap of unknown length
* 85276 86349: contig of 1074 bp in length
* 86350 86450: gap of unknown length
* 86450 87888: contig of 1439 bp in length
* 87889 89628: gap of unknown length
* 89629 89728: gap of unknown length
* 89729 91699: contig of 1971 bp in length

* 91700 91799: gap of unknown length
* 91800 92828: contig of 1029 bp in length
* 92829 92829: gap of unknown length
* 92929 94150: contig of 1222 bp in length
* 94151 94250: gap of unknown length
* 94251 95381: contig of 1131 bp in length
* 95382 95481: gap of unknown length
* 95482 96491: contig of 1010 bp in length
* 96492 96591: gap of unknown length
* 96592 97819: contig of 1228 bp in length
* 97820 97820: gap of unknown length
* 97920 99513: contig of 1594 bp in length
* 99514 99613: gap of unknown length
* 99614 100697: contig of 1084 bp in length
* 100698 100798: gap of unknown length
* 100799 102235: contig of 1438 bp in length
* 102236 102335: gap of unknown length
* 102336 103409: contig of 1074 bp in length
* 103410 103509: gap of unknown length
* 103510 104544: contig of 1035 bp in length
* 104545 104644: gap of unknown length
* 104645 105939: contig of 1295 bp in length.

FEATURES

Location/Qualifiers

Query Match 91.1%; Score 16.4; DB 2; Length 105939;
Best Local Similarity 94.4%; Pred. No. 94;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atactctgttcattgtaa 18

Db 97179 ATACTCTGTCATGTAA 97162

|||||

RESULT 8

AL359456_1

WPCOMMENT

Sequence split into 6 fragments LOCUS AL359456 Accession AL359456

Fragment Name	Begin	End
AL359456_0	1	110000
AL359456_1	100001	210000
AL359456_2	200001	310000
AL359456_3	300001	410000
AL359456_4	400001	510000
AL359456_5	500001	593964

Continuation (2 of 6) of AL359456 from base 100001 (AL359456 Homo sapiens chromosome

Query Match

Best Local Similarity 91.1%; Score 16.4; DB 2; Length 110000;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atactctgttcattgtaa 18

Db 13235 ATACTCTGTCATGTAA 13252

|||||

RESULT 9

AC091696/c

LOCUS

DEFINITION AC091696 137140 bp DNA linear HTG 19-MAY-2001
Felis catus clone RP86-117J4, WORKING DRAFT SEQUENCE, 8 unordered

AC091696

pieces

AC091696

AC091696.1 GI:14150418

HTG; HTGS_PHASE1; HTGS_DRAFT.

cat.

ORGANISM

Felis catus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.

1 (bases 1 to 137140)

Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W.,

Bouffard, G.G., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S.,

Guan, X., Gupta, J., Ho, S.-L., Idol, J.R., Karlins, E., Lee-Lin, S.-Q.,

Legaspi, R., Lim, M., Maduro, Q.L., Maduro, V.B., Masiello, C.,

Thu Aug 1 08:30:06 2002

Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Prasad, A.,
 Shevchenko, Y., Snyder, B., Stantrispop, S., Thomas, J.W., Thomas, P.J.,
 Tionsgon, E.E., Touchman, J.W., Tsurgon, C., Vogt, J.L., Walker, M.A.,
 Wetherby, K.D., Zhang, L.-H. and Green, E.D.
 NISC Comparative Sequencing Initiative
 Unpublished
 2 (bases 1 to 137140)
 Green, E.D.
 Direct Submission
 Submitted (19-MAY-2001) NIH Intramural Sequencing Center, 8717
 Government Circle, Gaithersburg, MD 20877, USA
 ----- Genome Center
 Center: NIH Intramural Sequencing Center
 Center code: NISC
 Web site: <http://www.nisc.nih.gov>
 Contact: nisc_mouse@nhgri.nih.gov
 ----- Project Information

 Center project name: awg
 Center clone name: 117J04
 ----- Summary Statistics
 Sequencing vector: plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 133760 bases at least Q40
 Consensus quality: 134423 bases at least Q30
 Consensus quality: 134718 bases at least Q20
 Insert size: 133000; agarose-fp
 Quality coverage: 11.63x in Q20 bases; agarose-fp
 Quality coverage: 11.33x in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently
 * consists of 8 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 2316: contig of 2316 bp in length
 * 2317 2416: gap of unknown length
 * 2417 13393: contig of 10977 bp in length
 * 13394 13493: gap of unknown length
 * 13494 20080: contig of 6587 bp in length
 * 20081 20180: gap of unknown length
 * 20181 31115: contig of 10935 bp in length
 * 31116 31215: gap of unknown length
 * 31216 47257: contig of 16042 bp in length
 * 47258 47357: gap of unknown length
 * 47358 72019: contig of 24662 bp in length
 * 72020 72119: gap of unknown length
 * 72120 102668: contig of 30549 bp in length
 * 102669 102768: gap of unknown length
 * 102769 137140: contig of 34372 bp in length.
 FEATURES
 Location/Qualifiers
 1. .137140
 /organism="Felis catus"
 /db_xref="taxon:9685"
 /clone="RP86-117J4"
 /clone_lib="RP86"
 1. .2316
 /note="assembly_fragment"
 2417. .13393
 /note="assembly_fragment"
 13494. .20080
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left
 20181. .31115
 /note="assembly_fragment"
 31216. .47257
 /note="assembly_fragment"
 clone_end:T7

vector_side:right"
 47358. .72019
 /note="assembly_fragment"
 72120. .102668
 /note="assembly_fragment"
 102769. .137140
 /note="assembly_fragment"
 BASE COUNT 37260 a 30755 c 29908 g 38508 t 709 others
 ORIGIN

Query Match 91.1%; Score 16.4; DB 2; Length 137140;
 Best Local Similarity 94.4%; Pred. No. 95;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atactctggctcatgttaa 18
 |||||
 Db 65046 ATACTCTGGTCATGTTGA 65029

RESULT 10
 AL591028/c 138716 bp DNA linear HTG 19-DEC-2001
 LOCUS Homo sapiens chromosome 10 clone RP13-143F22, *** SEQUENCING IN
 DEFINITION PROGRESS ***, 11 unordered pieces.
 AL591028
 ACCESSION AL591028.3 GI:117973956
 VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
 KEYWORDS human.
 SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Cranial; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 1 (sites)
 AUTHORS Burton, J.
 TITLE Direct Submission
 JOURNAL Submitted (18-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquery@sanger.ac.uk
 COMMENT On Dec 20, 2001 this sequence version replaced gi:14133148.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humquery@sanger.ac.uk
 ----- Project Information
 Center project name: b143F22
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: plasmid; 108752; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Consensus quality: 134354 bases at least Q40
 Consensus quality: 136114 bases at least Q30
 Consensus quality: 137107 bases at least Q20
 Insert size: 137716; sum-of-contigs
 Insert size: 151380; 17.0% error; agarose-fp
 Quality coverage: 5.19x in Q20 bases; sum-of-contigs Quality
 coverage: 4.77x in Q20 bases; agarose-fp

----- NOTE: This is a 'working draft' sequence. It currently
 * consists of 11 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 2424: contig of 2424 bp in length
 * 2425 2524: gap of 100 bp
 * 2525 24808: contig of 22284 bp in length
 * 24809 24908: gap of 100 bp
 * 24909 35362: contig of 10454 bp in length
 * 35363 35462: gap of 100 bp
 * 35463 43765: contig of 8303 bp in length

```

* 43766 43865: gap of 100 bp
* 43866 47699: contig of 3834 bp in length
* 47700 47799: gap of 100 bp
* 47800 78351: contig of 30552 bp in length
* 78352 78451: gap of 100 bp
* 78452 81539: contig of 3088 bp in length
* 81540 81639: gap of 100 bp
* 81640 88953: contig of 7314 bp in length
* 88954 89053: gap of 100 bp
* 89054 95324: contig of 6271 bp in length
* 95325 95424: gap of 100 bp
* 95425 98902: contig of 3478 bp in length
* 98903 99002: gap of 100 bp
* 99003 138716: contig of 39714 bp in length.
FEATURES
  source
    1. .138716
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="10"
      /clone="Rp13-143F22"
      /clone_lib="RPCI-13.1"
      1. .2424
        /note="assembly_fragment:01914
        fragment_chain:1"
        2525. .24808
          /note="assembly_fragment:02177
          fragment_chain:1"
          24909. .35362
            /note="assembly_fragment:01947
            fragment_chain:1"
            35463. .43765
              /note="assembly_fragment:00801
              fragment_chain:1"
              43866. .47699
                /note="assembly_fragment:01909
                fragment_chain:1"
                47800. .78351
                  /note="assembly_fragment:00910
                  fragment_chain:2"
                  78452. .81539
                    /note="assembly_fragment:02178
                    fragment_chain:2"
                    81640. .88953
                      /note="assembly_fragment:00075
                      fragment_chain:2"
                      89054. .95324
                        /note="assembly_fragment:01879
                        fragment_chain:2"
                        95425. .98902
                          /note="assembly_fragment:01238"
                          99003. .138716
                            /note="assembly_fragment:01608"
                            1000 others
BASE COUNT 45713 a 26430 c 25963 g 39610 t
ORIGIN

Query Match 91.1%; Score 16.4; DB 2; Length 138716;
Best Local Similarity 94.4%; Pred. No. 95;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atactctggtcatgtaa 18
||||| |||||||
Db 55073 ATACACTGGTCATGTAA 55056

RESULT 11
AC097356/c
LOCUS
DEFINITION Homo sapiens chromosome 3 clone RP11-107K23, WORKING DRAFT
ACCESSION AC097356
VERSION AC097356.1
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 147267)
Kaul, R.K., Olson, M.V., Raymond, C., Clendenning, J. and Haugen, E.D.
Direct Submission
Unpublished
2 (bases 1 to 147267)
Kaul, R.K., Olson, M.V., Raymond, C., Clendenning, J. and Haugen, E.D.
Direct Submission
Submitted (16-OCT-2001) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
On Oct 16, 2001 this sequence version replaced gi:7121074.
----- Genome Center
Center: University of Washington Genome Center
Center Code: UWGC
Web site: http://www.genome.washington.edu
Contact: uwgchtgs@u.washington.edu
Drafting Center: BCM
----- Project Information
Center project name: chr-3
Center clone name: RP11-107K23 (bc0228)
----- Summary Statistics
Sequencing vector: unknown; 65% of reads
Chemistry: Dye-terminator ET; 66% of reads
Chemistry: Dye-terminator Big Dye; 34% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 145614 bases at least Q40
Consensus quality: 146445 bases at least Q30
Consensus quality: 146882 bases at least Q20
Insert size: 147067; sum-of-contigs
Quality coverage: 9.4x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 8475: contig of 8475 bp in length
* 8476 8575: gap of unknown length
* 8576 47158: contig of 38583 bp in length
* 47159 47258: gap of unknown length
* 47259 147267: contig of 100009 bp in length.
FEATURES
  source
    1. .147267
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="3"
      /clone="RP11-107K23"
      /clone_lib="RPCI human BAC library 11"
      1. .8475
        /note="assembly_name:Contig12"
        8576. .47158
          /note="assembly_name:Contig13"
          47259. .147267
            /note="assembly_name:Contig14"
            210 others
BASE COUNT 45465 a 29529 c 28742 g 43321 t
ORIGIN

Query Match 91.1%; Score 16.4; DB 2; Length 147267;
Best Local Similarity 94.4%; Pred. No. 96;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atactctggtcatgtaa 18
||||| |||||||
Db 49216 ATACTCTGTCATGTAA 49199

```

RESULT 12

AC020698 148418 bp DNA linear PRI 09-MAY-2001
 LOCUS Homo sapiens BAC clone RP11-45F23 from 5, complete sequence.
 DEFINITION AC020698
 ACCESSION AC020698
 VERSION AC020698.4 GI:11120934
 KEYWORDS HTG.

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 148418)
 AUTHORS Sulston, J.E. and Waterston, R.
 TITLE Toward a complete human genome sequence
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
 MEDLINE 99063792

REFERENCE 2 (bases 1 to 148418)
 AUTHORS Nguyen, C., Drone, K., Hawkins, M. and Ureta, M.
 TITLE The sequence of Homo sapiens BAC clone RP11-45F23
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 148418)
 AUTHORS Waterston, R.H.

TITLE Direct Submission
 JOURNAL Submitted (08-JAN-2000) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

REFERENCE 4 (bases 1 to 148418)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (08-NOV-2000) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

REFERENCE 5 (bases 1 to 148418)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (09-MAY-2001) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Nov 8, 2000 this sequence version replaced gi:7630812.

----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc
 Contact: sapiens@watson.wustl.edu
 ----- Summary Statistics
 ----- Center project name: H_NH0045F23

NOTICE: This sequence may not represent the entire insert of this
 clone. It may be shorter because we only sequence overlapping
 clone sections once, or longer because we provide a small overlap
 between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate
 chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by sequence
 from more than one subclone; and the assembly was confirmed by
 restriction digest.

MAPPING INFORMATION:
 Mapping information for this clone was provided by Dr. John D.
 McPherson, Department of Genetics, Washington University, St. Louis
 MO. For additional information about the map position of this
 sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
 The RP11 human BAC library was made from the blood of one male
 donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E.,
 Tatenoe, M., Catanese, J.J. and de Jong, P.J. (1998) An improved
 approach for construction of bacterial artificial chromosome
 libraries. Genomics 51:1-8. The clone may be obtained either from
 Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong

and coworkers at the Roswell Park Cancer Institute
 (http://bacpac.med.buffalo.edu)

VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:
 The clone sequenced to the left is RP11-462G22. Actual start of
 this clone is at base position 1 of RP11-45F23; actual end is at
 base position 148418 of RP11-45F23.

Location/Qualifiers

FEATURES	source	1..148418	/organism="Homo sapiens"
			/db_xref="taxon:9606"
			/chromosome="2"
			/map="5"
			/clone="RP11-45F23"
			/clone_lib="RPC1-11"
		1..434	
repeat_region			/rpt_family="L1"
		507..807	
repeat_region			/rpt_family="L1"
		807..2585	
repeat_region			/rpt_family="L1"
		2616..3020	
repeat_region			/rpt_family="L1"
		3062..3250	
repeat_region			/rpt_family="L1"
		3251..3596	
repeat_region			/rpt_family="MER2_type"
		3597..3843	
repeat_region			/rpt_family="L1"
		9001..9137	
repeat_region			/rpt_family="Alu"
		11353..11406	
repeat_region			/rpt_family="ERVL"
		17191..17869	
repeat_region			/rpt_family="L1"
		19869..20177	
repeat_region			/rpt_family="Alu"
		21287..22345	
repeat_region			/rpt_family="L1"
		22727..22937	
repeat_region			/rpt_family="MER2_type"
		22933..23133	
repeat_region			/rpt_family="MER2_type"
		23355..23442	
repeat_region			/rpt_family="MIR"
		23485..23625	
repeat_region			/rpt_family="MER2_type"
		24107..24872	
repeat_region			/rpt_family="L1"
		24873..25320	
repeat_region			/rpt_family="ERV1"
		25372..25522	
misc_feature			/note="Similar to EST BE087288 (NID:98477683)"
		28473..28614	
repeat_region			/rpt_family="ERV1"
		28622..28736	
repeat_region			/rpt_family="ERV1"
		28790..28843	
repeat_region			/rpt_family="ERV1"
		28837..29212	
repeat_region			/rpt_family="ERV1"
		29215..29821	
repeat_region			/rpt_family="ERV1"
		29823..29972	
repeat_region			/rpt_family="ERV1"
		29974..30286	
repeat_region			/rpt_family="ERV1"
		30287..30737	
repeat_region			/rpt_family="ERV1"
		30745..31216	
repeat_region			/rpt_family="L1"
		31227..31275	
repeat_region			/rpt_family="MIR"

* arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1026: contig of 1026 bp in length
 1027 1126: gap of 100 bp
 1127 1306: contig of 180 bp in length
 1307 1406: gap of 100 bp
 1407 2680: contig of 1274 bp in length
 2681 2780: gap of 100 bp
 2781 4606: contig of 1826 bp in length
 4607 4706: gap of 100 bp
 4707 7010: contig of 2304 bp in length
 7011 7110: gap of 100 bp
 7111 10959: contig of 3849 bp in length
 10960 11059: gap of 100 bp
 11060 15310: contig of 4251 bp in length
 15311 15410: gap of 100 bp
 15411 20672: contig of 5262 bp in length
 20673 20772: gap of 100 bp
 20773 26614: contig of 5842 bp in length
 26615 26714: gap of 100 bp
 26715 32318: contig of 5604 bp in length
 32319 32418: gap of 100 bp
 32419 39439: contig of 7021 bp in length
 39440 39539: gap of 100 bp
 39540 47837: contig of 8298 bp in length
 47838 47937: gap of 100 bp
 47938 58126: contig of 10189 bp in length
 58127 58226: gap of 100 bp
 58227 66313: contig of 8087 bp in length
 66314 66413: gap of 100 bp
 66414 78356: contig of 11943 bp in length
 78357 78456: gap of 100 bp
 78457 90502: contig of 12046 bp in length
 90503 90602: gap of 100 bp
 90603 110197: contig of 19595 bp in length
 110198 110297: gap of 100 bp
 110298 132734: contig of 22437 bp in length
 132735 132834: gap of 100 bp
 132835 164235: contig of 31401 bp in length.

Key Location/Qualifiers

source 1..164235
 /chromosome="2"
 /db_xref="taxon:9606"
 /organism="Homo sapiens"
 /map="2"
 /clone="RP11-704A16"
 /clone.lib="RPC1-11 Human Male BAC"

1..1026
 /note="assembly_fragment"
 1127..1306
 /note="assembly_fragment clone_end:T7 vector_side:right"
 1407..2680
 /note="assembly_fragment"
 2781..4606
 /note="assembly_fragment"
 4707..7010
 /note="assembly_fragment"
 7111..10959
 /note="assembly_fragment"
 11060..15310
 /note="assembly_fragment"
 15411..20672
 /note="assembly_fragment"
 20773..26614
 /note="assembly_fragment"
 26715..32318
 /note="assembly_fragment"
 32419..39439
 /note="assembly_fragment"

FT misc_feature /note="assembly_fragment"
 FT 39540..47837
 FT /note="assembly_fragment"
 FT 47938..58126
 FT /note="assembly_fragment"
 FT 58227..66313
 FT /note="assembly_fragment clone_end:SP6 vector_side:left"
 FT 66414..78356
 FT /note="assembly_fragment"
 FT 78457..90502
 FT /note="assembly_fragment"
 FT 90603..110197
 FT /note="assembly_fragment"
 FT 110298..132734
 FT /note="assembly_fragment"
 FT 132835..164235
 FT /note="assembly_fragment"
 XX Sequence 164235 BP; 52554 A; 29032 C; 29120 G; 51728 T; 1801 other;
 SQ

Query Match 91.1%; Score 16.4; DB 30; Length 164235;
 Best Local Similarity 94.4%; Pred. No. 96;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 atactctgtcatgttaa 18
 ||| ||||| ||||| |||||
 DB 95389 ATAATCTGTCATGTAA 95372

RESULT 14
 AC108072/c 166922 bp DNA linear HTG 26-JAN-2002
 LOCUS Homo sapiens chromosome 2 clone RP11-704A16, WORKING DRAFT
 DEFINITION SEQUENCE, 2 unordered pieces.
 ACCESSION AC108072 AC036175
 VERSION AC108072.2 GI:18376983
 KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Waterston, R.H.
 The sequence of Homo sapiens clone
 Unpublished
 2 (bases 1 to 166922)
 Waterston, R.H.
 Direct Submission
 Submitted (24-JAN-2002) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 On Jan 26, 2002 this sequence version replaced gi:18308946.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 AUTHORS
 TITLE
 JOURNAL

COMMENT

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc/index.shtml
 Contact: submissions@wustl.edu
 ----- Project Information -----
 Center project name: H_NH0704A16
 Drafting center: WIBR
 ----- Summary Statistics -----
 Sequencing vector: M13; 43%
 Chemistry: Dye-terminator; 57%
 Chemistry: Dye-terminator; 0% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 166298 bases at least Q40
 Consensus quality: 166490 bases at least Q50
 Consensus quality: 166729 bases at least Q20
 Insert size: 168000; agarose-fp

Insert size: 166822; sum-of-contigs
Quality coverage: 11.20 in Q20 bases; agarose-fp
Quality coverage: 11.28 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1113: contig of 1113 bp in length
* 1114 1213: gap of unknown length
* 1214 166922: contig of 165709 bp in length.
FEATURES
source
1. 166922
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone="RP11-704A16"
1. 1113
/note="assembly_name:Contig4"
1214. 166922
/note="assembly_name:Contig8
clone_end:SP6
vector_side:right"
BASE COUNT 54810 a 30046 c 29856 g 52110 t 100 others
ORIGIN

Query Match 91.1%; Score 16.4; DB 2; Length 166922;
Best Local Similarity 94.4%; Pred. No. 96;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 atactctggctatgtaa 18
|||||
Db 100989 ATAATCTGGTCATGTTAA 100972

RESULT 15
AL356053 176355 bp DNA linear PRI 05-SEP-2001
LOCUS Human DNA sequence from clone RP11-453G10 on chromosome 10,
DEFINITION complete sequence.
ACCESSION AL356053
VERSION AL356053.14 GI:15487168
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 176355)
Laird,G.
Direct Submission
Submitted (05-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerquest@sanger.ac.uk
On Sep 6, 2001 this sequence version replaced gi:14272282.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given

in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TrEMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr10
RP11-453G10 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6
IMPORTANT: This sequence is not the entire insert of clone RP11-453G10 It may be shorter because we sequence overlapping sections only once, except for a short overlap.
The true right end of clone RP11-453G10 is at 176355 in this sequence. The true right end of clone RP11-19501 is at 100 in this sequence.

FEATURES
source
1. 176355
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-453G10"
/clone_lib="RPCI-11.2"
6975. 7489
/note="match: GSS: Em:AQ730872"
6983. 7519
/note="match: GSS: Em:AQ272795"
30521. 30876
/note="match: GSS: Em:AQ210265"
complement(35673..36133)
/note="match: GSS: Em:AQ632919"
35866. 36208
/note="match: GSS: Em:AQ142257"
complement(77773..78285)
/note="match: GSS: Em:AQ149763"
complement(101941..102380)
/note="match: GSS: Em:AQ030914"
complement(118019..118511)
/note="match: GSS: Em:AQ628733"
complement(153629..153876)
/note="match: GSS: Em:AQ267194"
164173..164612
/note="match: GSS: Em:AQ133255"
171407..171956
/note="match: GSS: Em:AQ678987"
BASE COUNT 53311 a 33340 c 33627 g 56077 t
ORIGIN

Query Match 91.1%; Score 16.4; DB 9; Length 176355;
Best Local Similarity 94.4%; Pred. No. 97;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 atactctggctatgtaa 18
|||||
Db 124306 ATACACTGGTCATGTTAA 124323

Search completed: July 31, 2002, 14:02:35
Job time: 17597 sec

us-09-899-718a-8.rge

Thu Aug 1 08:30:06 2002


```

; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
;
; SEQ ID NO 8
; LENGTH: 230
; TYPE: DNA
; ORGANISM: Homo sapiens
HIS-09-437-457-8

```

Query Match	71.7%;	Score 17.2;	DB 4;	Length 230;
Best Local Similarity	86.4%;	Pred. No. 75;		
Matches 19: Conservative	0;	Mismatches	3;	Indels 0;
				Gaps 0;

Qy 2 cagccagttccacccccgtgcac 23
 | ||||| ||||| ||||| |
Db 42 CTGCCAGTTCACGCCCGTGCTC 21

```

3
RESULT
US-08-385-370-1
; Sequence 1, Application US/08385370
; Patent No. 5665585
; GENERAL INFORMATION:
; APPLICANT: Torkkeli, Tuula
; APPLICANT: Joutsjoki, Vesa
; APPLICANT: Torkkeli, Helena
; APPLICANT: Vainio, Arja
; APPLICANT: Fagerstrom, Richard
; APPLICANT: Aho, Sirpa
; APPLICANT: Korhola, Matti
; APPLICANT: Nevalainen, Helena
; TITLE OF INVENTION: Production of Glucoamylase P With High
; TITLE OF INVENTION: Limit Dextrinase Activity in Trichoderma
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Ave., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/385,370

```

FILING DATE: 05/08/2004
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA: US/08/104,853

APPLICANT'S NAME: MORRIS, C. 05/20/2007 23:47:02
 BILLING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Cimbala, Michele A.
 REGISTRATION NUMBER: 33,851
 REFERENCE/DOCKET NUMBER: 1050.0270004
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 371-2600
 TELEFAX: (202) 371-2540

TELEGRAMS: 12027 571 2510
TELEX: 248636 SSK
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1996 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both

Query Match 71.7%; Score 17.2; DB 1; Length 1996;
Best Local Similarity 86.4%; Pred. No. 76;
Matches 19; Conservative 0; Mismatches 3; Indels 0

2y 1 tcagccagtctccaccccgtgca 22
||| ||| ||| ||| ||| ||| |||
bb 1473 TCCGCCAGCTCCACCCCCGGCA 1494

RESULT 4
US-08-385-370-3
sequence 3 Application US/08385370

Sequence 37, Affiliation 067388383

Patent No. 5665585

GENERAL INFORMATION:

APPLICANT: Torkkeli, Tuula

APPLICANT: Joutsjoki, Vesa

APPLICANT: Torkkeli, Helena

APPLICANT: Vainio, Arja

APPLICANT: Fagerstrom, Richard

APPLICANT: Aho, Sirpa

APPLICANT: Korhola, Matti

APPLICANT: Nevalainen, Helena

TITLE OF INVENTION: Production of Glucoamylase P With High

TITLE OF INVENTION: Limit Dextrinase Activity in Trichoderma

NUMBER OF SEQUENCES: 7

NUMBER OF COPIES OF CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Ave., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005

```

1  ZIF: 20003
2  COMPUTER READABLE FORM:
3  MEDIUM TYPE: Floppy disk
4  COMPUTER: IBM PC compatible
5  OPERATING SYSTEM: PC-DOS/MS-DOS
6  SOFTWARE: Patent In Release #1.0, Version #1.25
7  CURRENT APPLICATION DATA:
8  PUBLICATION NUMBER: IIS/08/385, 370

```

```

; APPLICATION NUMBER: US/08/365,575
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/104,853
; FILING DATE:

```

FILED DATE: 02/01/2004
 ATTORNEY/AGENT INFORMATION:
 NAME: Cimbala, Michele A.
 REGISTRATION NUMBER: 33,851
 REFERENCE/DOCKET NUMBER: 1050.0270004
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 371-2600
 TELEFAX: (202) 371-2540
 TELEX: 248636 SSK

```

, INDEX: 2000000000
, INFORMATION FOR SEQ ID NO: 3:
,-----
, SEQUENCE CHARACTERISTICS:
,     LENGTH: 2745 base pairs
,     TYPE: nucleic acid
,     STRANDEDNESS: both
,     TOPOLOGY: both
,-----
, FEATURE:
,     NAME/KEY: CDS
,     LOCATION: join(295..520,
, 15-08-385-370-3

```

Query Match	71.7%;	Score 17.2;	DB 1;	Length 2745;
Best Local Similarity	86.4%;	Pred. No. 76;		
Matches 19: Conservative	0;	Mismatches 3;	Indels 0	

Qy 1 tcagccagtccaccccgtgca 22
 || ||||| ||||||||| |||
Dh 2040 tccggcaggctccaccggcgca 2061

RESULT 5
US-08-303-861-1
. Sequence 1. Application US/08303861

```

RESULT      6
US-09-060-410-3/c
; Sequence 3, Application US/09060410
; Patent No. 6165461
; GENERAL INFORMATION:
; APPLICANT: Cobb, Melanie
; APPLICANT: Hutchinson, Michele
; APPLICANT: Chen, Zhu
; APPLICANT: Berman, Kevin
; TITLE OF INVENTION: TAO PROTEIN KINASES AND METHODS OF USE
; TITLE OF INVENTION: THEREFOR
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:

```

Query Match	69.2%	Score 16.6;	DB 4;	Length 374;
Best Local Similarity	79.2%;	Pred. No. 1.3e+02;		
Matches 19;	Conservative 0;	Mismatches 5;	Indels 0;	Gaps 0;
QY	1	tcagccagttccaccccgtagcag	24	
Db	101	TCGGCCAGNTCCACGGCGGCCACG	78	

CLASSIFICATION: 800
APPLICATION NUMBER: PCT/US93/09340
FILING DATE: 30-SEP-1993
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Kitchell, Barbara S.
REGISTRATION NUMBER: 33,928
REFERENCE/DOCKET NUMBER: ARCD:152/KIT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (713) 789-2679
TELEX: 79-0924
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1458 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
US-08-468-793-1

Query Match 69.2%; Score 16.6; DB 4; Length 1458;
Best Local Similarity 82.6%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 tcagccaggttcaccccgtagc 23
Db 387 TTAGCCAGTTCACCTCCTAGC 365

RESULT 11
US-08-581-148C-15
Sequence 15, Application US/08581148C
Patent No. 6060644
GENERAL INFORMATION:
APPLICANT: Schnable, Patrick S.
APPLICANT: Robertson, Donald S.
APPLICANT: Hansen, Joel D.
APPLICANT: Nikolau, Basil J.
APPLICANT: Xu, Xiaojie
TITLE OF INVENTION: ISOLATION AND USE OF CUTICULAR LIPID
TITLE OF INVENTION: GENES
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Leydig, Voit & Mayer, Ltd.
STREET: Two Prudential plaza, Suite 4900
CITY: Chicago
STATE: IL
COUNTRY: US
ZIP: 60601-6780

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/581,148C
FILING DATE: 29-DEC-1995
CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:
NAME: Larcher, Carol
REGISTRATION NUMBER: 35243
REFERENCE/DOCKET NUMBER: 71380
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 616-5600
TELEFAX: (312) 616-5700
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 1903 base pairs
TYPE: nucleic acid
STRANDEDNESS: double

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1665
US-08-581-148C-15

Query Match 69.2%; Score 16.6; DB 3; Length 1903;
Best Local Similarity 82.6%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 tcagccaggttcaccccgtagc 23
Db 266 TCAGCCGTTACCACTCCTTGAC 288

RESULT 12
US-07-903-029-2/c
Sequence 2, Application US/07903029
Patent No. 5969097
GENERAL INFORMATION:
APPLICANT: Wiegand, Roger C.
APPLICANT: Currie, Mark C.
APPLICANT: Fok, Kam F.
TITLE OF INVENTION: Human Guanylin
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dennis A. Bennett, Monsanto Co., A3SG
STREET: 800 N. Lindbergh Blvd.
CITY: St. Louis
STATE: Missouri
COUNTRY: USA
ZIP: 63167

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/903,029
FILING DATE: 19920623
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Bennett, Dennis A.
REGISTRATION NUMBER: 34,547
REFERENCE/DOCKET NUMBER: 07-21(872)A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (314)694-5402
TELEFAX: (314)694-9009
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 589 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-07-903-029-2

Query Match 67.5%; Score 16.2; DB 2; Length 589;
Best Local Similarity 85.7%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 tcagccaggttcaccccgtagc 21
Db 500 TCACCCAGTTCCTCCCGGAC 480

RESULT 13
US-09-155-768-1/c
Sequence 1, Application US/09155768A
Patent No. 6162908

us-09-899-718a-7.rni

Thu Aug 1 08:30:05 2002

```
; GENERAL INFORMATION:
; APPLICANT: SEIKAGAKU CORPORATION
; TITLE OF INVENTION: POLYPEPTIDE OF HYALURONAN SYNTHASE OF
; TITLE OF INVENTION: HUMAN-ORIGIN AND DNA ENCODING THE SAME
; FILE REFERENCE: TOYAM37_001APC
; CURRENT APPLICATION NUMBER: US/09/155,768A
; CURRENT FILING DATE: 1998-10-05
; EARLIER APPLICATION NUMBER: JP 8-084326
; EARLIER FILING DATE: 1996-04-05
; EARLIER APPLICATION NUMBER: JP 8-109663
; EARLIER FILING DATE: 1996-04-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 2117
; TYPE: DNA
; ORGANISM: HUMAN
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (149)...(1777)
US-155-768-1
```

```
Query Match      67.5%; Score 16.2; DB 4; Length 2117;
Best Local Similarity 85.7%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      4 gccagttccaccccgtagc 24
        ||||| ||||| || |||||
DB      1374 GCCAGTGGCAGCCCTGCACG 1354
```

```
RESULT 14
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2
```

```
Query Match      67.5%; Score 16.2; DB 4; Length 4403765;
Best Local Similarity 90.0%; Pred. No. 54;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      2 cagccagttccaccccgtagc 21
        ||||| ||||| || |||||
DB      2098669 CACCCAGTTCACCCCGGTGC 2098650
```

```
RESULT 15
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
```

```
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1
```

```
Query Match      67.5%; Score 16.2; DB 4; Length 4411529;
Best Local Similarity 90.0%; Pred. No. 54;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      2 cagccagttccaccccgtagc 21
        || ||||| ||||| |||||
DB      2101479 CACCCAGTTCACCCCGGTGC 2101460
```

```
Search completed: July 31, 2002, 12:26:27
Job time: 12010 sec
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 12:12:09 ; Search time 5855.71 Seconds
(without alignments)
55.318 Million cell updates/sec

Title: US-09-899-718A-7
Perfect score: 24
Sequence: 1 tcagcagttccaccctgcacg 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_hic.*
9: gb_est1.*
10: gb_est2.*
11: gb_hic.*
12: gb_gss.*
13: em_gss_hum.*
14: em_gss_inv.*
15: em_gss_pln.*
16: em_gss_vrt.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	19.4	80.8	836	10	BG444193
C 2	19.2	80.0	808	12	CNS04FCX
C 3	18.8	78.3	420	10	C19505
C 4	18.8	78.3	453	9	AU065405
C 5	18.8	78.3	460	9	AU065404
C 6	18.2	75.8	306	10	B1510068
C 7	18.2	75.8	426	10	BG981373
C 8	18.2	75.8	528	9	A1162491
C 9	18.2	75.8	559	10	BF191609
C 10	18.2	75.8	687	10	BG827582
C 11	18.2	75.8	702	10	B1329586
C 12	18.2	75.8	765	12	BH045197
C 13	18.2	75.8	907	10	BG104185
C 14	18.2	75.8	1200	11	AK009013
C 15	17.8	74.2	64	12	AZ875573
C 16	17.8	74.2	449	10	BF739541
C 17	17.8	74.2	642	12	AZ831095

C 18	17.8	74.2	768	10	BG962250
C 19	17.8	74.2	1021	12	CNS03RVF
C 20	17.6	73.3	276	10	BF527986
C 21	17.6	73.3	286	9	A1104620
C 22	17.6	73.3	325	10	BF415383
C 23	17.6	73.3	327	9	A1639268
C 24	17.6	73.3	340	9	A1414355
C 25	17.6	73.3	346	9	AV012778
C 26	17.6	73.3	373	12	AZ790364
C 27	17.6	73.3	385	12	AZ790223
C 28	17.6	73.3	386	10	BF407002
C 29	17.6	73.3	393	9	AW523973
C 30	17.6	73.3	400	10	BI995885
C 31	17.6	73.3	413	9	AW433934
C 32	17.6	73.3	416	9	AL502941
C 33	17.6	73.3	460	9	AW434970
C 34	17.6	73.3	487	9	BB857929
C 35	17.6	73.3	519	10	BM426700
C 36	17.6	73.3	522	10	BI993715
C 37	17.6	73.3	523	9	AA816166
C 38	17.6	73.3	524	9	AI391290
C 39	17.6	73.3	542	9	AI454923
C 40	17.6	73.3	569	10	BG373840
C 41	17.6	73.3	579	9	AV932786
C 42	17.6	73.3	716	10	BG843380
C 43	17.6	73.3	733	10	BE566538
C 44	17.6	73.3	769	10	BF627856
C 45	17.6	73.3	806	10	BG334627

ALIGNMENTS

RESULT 1

BG444193/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BG444193 836 bp mRNA linear EST 15-MAR-2001
GA_Ea0023J06f Gossypium arboreum 7-10 dpa fiber library Gossypium
arboreum cDNA clone GA_Ea0023J06f, mRNA sequence.
BG444193
BG444193.1 GI:13353845
EST.
Gossypium arboreum.
Gossypium arboreum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
1 (bases 1 to 836)
Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry
D., Wood,T.C., Leslie,A. and Wilkins,T.A.
An integrated analysis of the genetics, development, and evolution
of the cotton fiber
Unpublished (2000)
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAATACGACTCTACTATAGGG
High quality sequence start: 2
High quality sequence stop: 826.
Location/Qualifiers
1. 836
/organism="Gossypium arboreum"
/strain="ARA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA_Ea0023J06f"
/tissue_type="Fibers isolated from bolls harvested 7-10
dpa"
/lab_host="E. coli"

FEATURES

source

1. 836
/organism="Gossypium arboreum"
/strain="ARA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA_Ea0023J06f"
/tissue_type="Fibers isolated from bolls harvested 7-10
dpa"
/lab_host="E. coli"

Thu Aug 1 08:30:06 2002

```

C19505
LOCUS          C19505 420 bp mRNA linear EST 24-OCT-1996
DEFINITION    C19505 Rice panicle at ripening stage Oryza sativa cDNA clone
               E10525_1A, mRNA sequence.
ACCESSION     C19505.1 GI:1631776
VERSION       C19505.1
KEYWORDS      EST.
SOURCE        Oryza sativa.
ORGANISM      Oryza sativa.
REFERENCE     1 (bases 1 to 420)
AUTHORS       Sasaki,T.
TITLE         Rice cDNA from panicle at ripening stage
JOURNAL       Unpublished (1996)
COMMENT       Contact: Takuji Sasaki
               National Institute of Agrobiological Resources
               Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
               305-8602, Japan
               Tel: 81-298-38-7461
               Fax: 81-298-38-7468
               Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/.

FEATURES             source
   1..420
   /organism="Oryza sativa"
   /strain="Nipponbare"
   /db_xref="taxon:4530"
   /clone="E10525_1A"
   /clone.lib="Rice panicle at ripening stage"
   /dev_stage="ripening stage"
   /note="Organ: panicle; Rice cDNA from panicle at ripening
   stage"

BASE COUNT  92 a 132 c 121 g 71 t 4 others
ORIGIN

Query Match      78.3%; Score 18.8; DB 10; Length 420;
Best Local Similarity 90.9%; Pred. No. 1.3e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY  1 tcagccagttccaccctgcac 22
    ||| ||||| ||||| ||||| |||
Db   323 TCGCCAGTTCACCCCGTGAA 344

RESULT  4
LOCUS    AU065405 453 bp mRNA linear EST 20-MAY-1999
DEFINITION AU065405 Rice root Oryza sativa cDNA clone R3592_1A, mRNA sequence.
ACCESSION AU065405
VERSION    AU065405.1 GI:4881106
KEYWORDS   EST.
SOURCE      Oryza sativa.
ORGANISM    Oryza sativa.
REFERENCE   1 (bases 1 to 453)
AUTHORS     Minobe,Y. and Sasaki,T.
TITLE       Rice cDNA from root
JOURNAL     Unpublished (1995)
COMMENT     Contact: Takuji Sasaki
               National Institute of Agrobiological Resources
               Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
               305-8602, Japan
               Tel: 81-298-38-7441
               Fax: 81-298-38-7468
               Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
               PROJECT "RGP"

FEATURES             source
   1..453
   /organism="Oryza sativa"

CNS04FCX
LOCUS          CNS04FCX 808 bp DNA linear GSS 21-MAY-2000
DEFINITION    CNS04FCX Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
               106E04 of library G from Tetraodon nigroviridis, genomic survey
               sequence.
ACCESSION     AL288186
VERSION       AL288186.1 GI:8026711
KEYWORDS      GSS: genome survey sequence.
SOURCE        Tetraodon nigroviridis.
ORGANISM      Tetraodon nigroviridis.
REFERENCE     1 (bases 1 to 808)
AUTHORS       Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
               Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
               Weissenbach,J.
TITLE         Characterization and repeat analysis of the compact genome of the
               freshwater pufferfish Tetraodon nigroviridis
JOURNAL       Unpublished
REFERENCE     2 (bases 1 to 808)
AUTHORS       Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
               Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
               Saurin,W. and Weissenbach,J.
TITLE         Human gene number estimate provided by genome wide analysis using
               Tetraodon nigroviridis DNA sequence
JOURNAL       Unpublished
REFERENCE     3 (bases 1 to 808)
AUTHORS       Genoscope.
TITLE         Direct Submission
JOURNAL       Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT       This sequence is a single read and was generated as part of a large
               scale clone-end sequencing project of the Tetraodon nigroviridis
               genome. For more information, please take a look at
               http://www.genoscope.cns.fr/Tetraodon.

FEATURES             source
   1..808
   /organism="Tetraodon nigroviridis"
   /db_xref="taxon:99883"
   /clone="106E04"
   /clone.lib="G"
   /note="Genoscope sequence ID : C0BG106BC02SP1-end :
   PUC-Ori"

BASE COUNT  149 a 222 c 246 g 185 t 6 others
ORIGIN

Query Match      80.0%; Score 19.2; DB 12; Length 808;
Best Local Similarity 87.5%; Pred. No. 1e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY  1 tcagccagttccaccctgcac 24
    ||| ||||| ||||| ||||| |||
Db   225 TCAGCCAGTTCACCCGCTGCAG 202

RESULT  3

```

```

/strain="Nipponbare, sub_species Japonica"
/db_xref="taxon:4530"
/clone="R3592_1A"
/clone_lib="Rice root"
/note="Prepared from seedling root."
BASE COUNT      103 a 136 c 126 g 84 t 4 others
ORIGIN

```

```

Query Match      78.3%; Score 18.8; DB 9; Length 453;
Best Local Similarity 90.9%; Pred. No. 1.3e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tcagccagttccaccctgtgca 22
   ||| ||||| ||||| ||||| |||
Db 325 TCGGCCAGTTCACCCCGTGAA 346

```

```

RESULT 5
AU065404
LOCUS
DEFINITION AU065404 Rice root Oryza sativa cdna clone R3590_1A, mRNA sequence.
ACCESSION AU065404
VERSION AU065404.1 GI:4881105
KEYWORDS EST.
SOURCE
ORGANISM Oryza sativa.

```

```

REFERENCE
AUTHORS Minobe, Y. and Sasaki, T.
TITLE Rice cdna from root
JOURNAL Unpublished (1995)
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsasaki@abr.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/
PROJECT = RGP.

```

```

FEATURES
source
Location/Qualifiers
1..460
/organism="Oryza sativa"
/strain="Nipponbare, sub_species Japonica"
/db_xref="taxon:4530"
/clone="R3590_1A"
/clone_lib="Rice root"
/note="Prepared from seedling root."
BASE COUNT      103 a 145 c 128 g 80 t 4 others
ORIGIN

```

```

Query Match      78.3%; Score 18.8; DB 9; Length 460;
Best Local Similarity 90.9%; Pred. No. 1.3e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tcagccagttccaccctgtgca 22
   ||| ||||| ||||| ||||| |||
Db 328 TCGGCCAGTTCACCCCGTGAA 349

```

```

RESULT 6
BI510068
LOCUS
DEFINITION BI510068 BBI70026B10G05 Bee Brain Normalized/Subtracted Library, BBI7 Apis mellifera cdna clone BBI70026B10G05 5', mRNA sequence.
ACCESSION BI510068
VERSION BI510068.1 GI:15360442
KEYWORDS EST.
SOURCE honeybee.
ORGANISM Apis mellifera

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea; Apidae; Apis.

1 (bases 1 to 306)
Whitfield, C.W., Soares, B., Robertson, H.M., Pardinas, J., Liu, L., Smoller, D. and Robinson, G.E.
An Expressed Sequence Tag Resource for Studies of Brain and Behavior in the Honey Bee

Unpublished (2001)
Contact: Gene E. Robinson
Department of Entomology
University of Illinois
505 S. Goodwin Ave., Urbana, IL 61801, USA
Tel: 217 265 0309
Fax: 217 244 3499
Email: generobi@life.uiuc.edu

This research was funded by the University of Illinois Critical Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation Award in Functional Genomics to G.E. Robinson and an NSF Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
PCR PRIMERS

FORWARD: TANTACGACTCATTATAGGG

BACKWARD: ATTAACCCCTCACTAAAG

Insert Length: 306 Std Error: 0.00

Plate: BBI70026B10 row: G column: 05

Seq primer: AGCGGATAACAATTTCACACAGGA

High quality sequence stop: 306.

Location/Qualifiers

1..306

/organism="Apis mellifera"
/strain="mixed strains of European bees, predominantly A.m. ligustica"
/db_xref="taxon:7460"
/clone_lib="BBI70026B10G05"
/clone="BBI70026B10G05"
/sex="female"
/tissue_type="brain"
/dev_stage="adult worker honey bee"
/lab_host="DH10B"

/note="Organ: brain; Vector: pT73-Pac; Site_1: EcoRI; Site_2: NotI; This BBI7 cdna library was generated by subtraction of the BBI6 library with 4000 previously sequenced clones. The BBI6 library was contributed by the Soares laboratory and it was constructed and normalized by the Soares laboratory by Ronaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome Research 6(9): 791-806. RNA was prepared from dissected brains of adult worker bees of various ages and various behavioral groups."

BASE COUNT 94 a 67 c 59 g 86 t

ORIGIN

Query Match 75.8%; Score 18.2; DB 10; Length 306;
Best Local Similarity 87.0%; Pred. No. 2.2e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

Qy 1 tcagccagttccaccctgtgca 23
   ||| ||||| ||||| ||||| |||
Db 256 TCGGCCAGTTCACCCCGTGCTC 278

```

```

RESULT 7
BG981373/c
LOCUS
DEFINITION BG981373 CM4-CN0089-070201-746-e01 CN0089 Homo sapiens cdna, mRNA sequence.
ACCESSION BG981373
VERSION BG981373.1 GI:14384108
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

```

```

REFERENCE
1 (bases 1 to 426)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

AUTHORS
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.D. and Simpson, A.J.G.

TITLE
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE
20202663

ABSTRACT
Contact: Simpson A.J.G.

Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/IICR Human Cancer Genome
Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM4&t2=CM4-CN0089-070201-746-e01&t3=2001-02-07&t4=1>)
Seq primer: puc 18 forward
High quality sequence start: 54
High quality sequence stop: 426.

FEATURES	source	Location/Qualifiers
	1. 426	
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/clone_lib="CN0089"
		/dev_stage="Adult"
		/note="Organ: colon.normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT	85 a 131 c 103 g 107 t	
ORIGIN		

```

Query Match          75.8%;   Score 18.2; DB 10; Length 426;
Best Local Similarity 87.0%;   Pred. No. 2.3e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

      1 tcagccagtgtccaccccgtgcac 23
        ||||| ||||||||| |||
      299 TCAGACAGTTCGACCCCATACAC 277

RESULT 8
AII162491 linear EST 03-DEC-1998
LOCUS AII162491 528 bp mRNA
DEFINITION A018P38U Hybrid assay plasmid library Populus tremula x Populus
            A018P38U hybrid cDNA 5' mRNA sequence.

```

ACCESSION	A1162491
VERSION	A1162491.1
KEYWORDS	GI:3853776
SOURCE	EST.
ORGANISM	Populus tremula x Populus tremuloides. Populus tremula x Populus tremuloides Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Malpighiales; Salicaceae; Populus. 1 (bases 1 to 528)
REFERENCE	Sterky,F.F., Regan,S., Karlsson,J., Hertzberg,M., Rohde,A., Holmberg, A., Anini,B., Bhale Rao,R., Larsson,M., Villarreal,R., Van Montagu, M., Sandberg,G., Olsson,O., Teeri,T.T., Boerjan,W., Gustarsson,P., Uhlen,M., Sundberg,B. and Lundberg,J. Gene discovery in the wood-forming tissues of poplar: Analysis of 5,692 expressed sequence tags (22), 13330-13335 (1998) Proc. Natl. Acad. Sci. U.S.A.
TITLE	95 (22), 13330-13335 (1998)
JOURNAL	99007314
MEDLINE	99007314

Contact: Sterky, F
 Department of Biotechnology
 Royal Institute of Technology (KTH)
 Teknilingen 34, S-100 44 STOCKHOLM, Sweden
 Tel.: +46 8 790 8287
 Fax: +46 8 24 54 52
 Email: fredrik@biochem.kth.se
 PCR Primers
 FORWARD: AAAGGGGATGTGTCAGCGG
 BACKWARD: GCTTCGGCTCGTATGTTGTG
 Seq primer: CGTTGTAAGAGCGCCAG
 high quality sequence stop: 528.

```

FEATURES
  source
    Location/Qualifiers
      1..528
        /organism="Populus tremula x Populus tremuloides"
        /db_xref="taxon:47664"
        /clone_lib="Hybrid aspen plasmid library"
        /tissue_type="Cambial region"
        /dev_stage="1.5 m actively growing tree"
        /lab_host="E.coli"
        /note="Vector: pBluescript SK; Site.1: SalI; Site.2: NotI;
        Cambial region tissues, including developing xylem, the
        meristematic cambial zone and the developing and mature
        phloem, was harvested from 1.5 m actively growing trees.
        cDNA was prepared and cloned into lambda gt2a. DNA was
        isolated and subcloned into pBluescript SK using SalI and
        NotI restriction enzymes."
      116 a 174 c 111 g 124 t 3 others

```

Query Match	75.8%	Score 18.2	DB 9	Length 528
Best Local Similarity	87.0%	Pred. No. 2.3e+03		
Matches	20	Conservative 0	Mismatches 3	Indels 0
Gaps	0			
QY	1	tcagccaggttccaccccggtgcac	23	
Db	398	TCAGCCAGTTCGCCCGGGTGCAC	420	
RESULT	9			
BF191609		559 bp	mRNA	linear
LOCUS	BF191609		5' mRNA sequence	
				EST 02-NOV-2000

DEFINITION	239242 MARC 2P1G SUS Scrofa CDNA 3', 5' mRNA Sequence.
ACCESSION	BF191609
VERSION	BF191609.1 Gi:11074978
KEYWORDS	EST.
SOURCE	pig.
ORGANISM	Sus scrofa
REFERENCE	Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
AUTHORS	1 (bases 1 to 559) Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W., and Keele,J.W. of two pooled tissue normalized cDNA libraries for design and use of two pooled tissue normalized cDNA libraries for EST discovery and use in swine
TITLE	EST discovered (2000)
JOURNAL	Unpublished (2000)
COMMENT	Contact: Smith TPL USDA, ARS, US Meat Animal Research Center PO Box 166, Clay Center, NE 68933-0166, USA Tel: 402 762 4366 Fax: 402 762 4390 Email: smith@email.marc.usda.gov Single pass sequencing. Bases called and alt_trimmed with phred v.0.980904.e. Vector identified by cross_match with the -minscore 18 and -mismatch 12 options. PCR Primers FORWARD: AGGAACACGCTATGACCAT BACKWARD: GTTTTCCGAGTCACGAGC Plate: 67 row: B column: 7 Seq primer: ATTAGTGACACATAG. Location/Qualifiers

FEATURES

SOURCE

1..559
 /organism="Sus scrofa"
 /db_xref="taxon:9823"
 /clone_lib="MARC 2PIC"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
 Library made from pooled tissue from testis, ovary,
 endometrium, hypothalamus, pituitary, and placenta."
 BASE COUNT 117 a 181 c 152 g 109 t
 ORIGIN

Query Match 75.8%; Score 18.2; DB 10; Length 559;
 Best Local Similarity 87.0%; Pred. No. 2.4e+03;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 tcagccaggttcaccccggtgcac 23
 ||||| ||||| ||||| |||||
 Db 269 TCAGCCAGATCCACCCATGCAC 291

RESULT 10
 BG827582/c
 LOCUS
 DEFINITION 602748468F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4901416 5',
 mRNA sequence.
 ACCESSION BG827582
 VERSION BG827582.1 GI:14175169
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 687)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLCMI796 row: g column: 17
 High quality sequence start: 4
 High quality sequence stop: 551.
 Location/Qualifiers
 1..687

FEATURES

source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4901416"
 /clone_lib="NIH_MGC_17"
 /tissue_type="rhabdomyosarcoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;
 Site_2: XhoI; cDNA made by oligo-dt priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCAGAG(G). Size-selected >500bp
 for average insert size 1.8kb. Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."
 BASE COUNT 162 a 178 c 224 g 123 t
 ORIGIN

Query Match 75.8%; Score 18.2; DB 10; Length 687;
 Best Local Similarity 87.0%; Pred. No. 2.4e+03;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 tcagccaggttcaccccggtgcac 23
 ||||| ||||| ||||| |||||
 Db 136 TCAGCCAGCTCCACCGGTGCAC 114

RESULT 11
 BI329586/c
 LOCUS
 DEFINITION 602984002F1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:5136697 5',
 mRNA sequence.
 ACCESSION BI329586
 VERSION BI329586.1 GI:15014243
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 702)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM11336 row: c column: 02
 High quality sequence stop: 533.
 Location/Qualifiers
 1..702

FEATURES source
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:5136697"
 /clone_lib="NCI_CGAP_L19"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 1.9 kb. Constructed by Life
 Technologies. Note: this is a NCI_CGAP Library."
 BASE COUNT 130 a 237 c 204 g 131 t
 ORIGIN

Query Match 75.8%; Score 18.2; DB 10; Length 702;
 Best Local Similarity 87.0%; Pred. No. 2.4e+03;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 tcagccaggttcaccccggtgcac 23
 ||||| ||||| ||||| |||||
 Db 644 TCAGCCAGTCCACCCCTGTAC 622

RESULT 12
 BH045197/c
 LOCUS
 DEFINITION BH045197
 ACCESSION BH045197.1 GI:14829258
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 765)
 AUTHORS Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akınret, B., Levins, M.,
 Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregorgis, E.,

Query Match 75.8%; Score 18.2; DB 10; Length 765;
 Best Local Similarity 87.0%; Pred. No. 2.4e+03;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Thu Aug 1 08:30:06 2002

Russell, D., de Jong, P. and Fraser, C.M.
 Mouse BAC End Sequences from Library RPCI-24
 Unpublished (1999)
 Other GSSs: RPCI-24-301P3.TJ
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-24. For BAC
 library availability, please contact Pieter de Jong
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC
 Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end
 page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 301 row: P column: 3
 Seq primer: T7
 Class: BAC ends.

FEATURES

Location/Qualifiers
 1. .765
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-24-301P3"
 /clone_lib="RPCI-24"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /note="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI;
 RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
 library was cloned in the pTARBAC1 cloning vector at the
 BamHI sites using MboI partially digested male C57BL/6J
 DNA."

BASE COUNT 198 a 189 c 209 g 169 t

Query Match 75.8%; Score 18.2; DB 12; Length 765;
 Best Local Similarity 87.0%; Pred. No. 2.5e+03;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 tcagccaggttcaccccggtgcac 23
 | ||||| ||||| ||||| |||||
 Db 325 TGAGCCAGTTCACGCCCTGCAC 303

RESULT 13
 BG104185/c
 LOCUS
 DEFINITION
 602310751F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:4422900 5',
 mRNA sequence.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

Human.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

1 (bases 1 to 907)
 NIH-MGC <http://mgs.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabs-r@mail.nih.gov
 Tissue Procurement: ATCC/DCTD/DTP
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: LLCMI218 row: e column: 13
 High quality sequence stop: 710.
 Location/Qualifiers

FEATURES

source

1. .907
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4422900"
 /clone_lib="NIH_MGC_20"
 /tissue_type="melanotic melanoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(G). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and superscript II RT (Life Technologies)."
 BASE COUNT 146 a 275 c 339 g 147 t
 ORIGIN

Query Match 75.8%; Score 18.2; DB 10; Length 907;
 Best Local Similarity 87.0%; Pred. No. 2.5e+03;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 cagccagttccaccccggtgcac 24
 | ||||| ||||| ||||| |||||
 Db 666 CAGCCAGTCCACGCCCTGCAGG 644

RESULT 14
 AK009013/c
 LOCUS
 DEFINITION

AK009013 1200 bp mRNA linear HTC 19-JAN-2002
 Mus musculus adult male tongue cDNA, RIKEN full-length enriched
 library, clone:2300002M23:hypothetical protein, full insert
 sequence.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE

ORGANISM

Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED

1 (sites)
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 99279253
 10349636

REFERENCE

2 (sites)
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

4 (sites)
 The RIKEN Genome Exploration Research Group Phase II Team and the
 FANTOM Consortium.

us-09-899-718a-7.rst

Thu Aug 1 08:30:06 2002


```

XX PS Claim 7; Fig 1; 65pp; English.
XX
CC The sequence is derived from 3 overlapping clones, pRH3-1, pRH4-23,
CC and pRH4-31. (Deposited as ATCC 67885, 67886, and 67887 resp.)
CC The clones were isolated from a cDNA library in the lambda Zap
CC vector prep'd. from mRNA obtd. from newborn rat hearts using rat
CC brain II cDNA probe. The isolated DNA can be used to screen a
CC similar human derived cardiac cDNA library for the corresponding
CC human gene. Proteins produced by expression of the DNA have
CC diagnostic therapeutic, and prognostic applications.
XX
XX SQ Sequence 7555 BP; 1576 A; 2314 C; 2101 G; 1564 T; 0 other;

Query Match 75.8%; Score 18.2; DB 11; Length 7555;
Best Local Similarity 87.0%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 tcagccagttccaccgccgtgcac 23
   ||||| ||||| ||||| |||||
Db 539 tcagcccttcaccgccgtgcgc 561

RESULT 2
AAQ81328
ID AAQ81328 standard; cDNA; 7555 BP.
XX
AC AAQ81328;
XX
DT 05-AUG-1995 (first entry)
XX
DE Cardiac sodium channel protein coding sequence.
XX
KW Sodium channel protein; ds; therapeutic; diagnostic; prognostic;
KW antiarrhythmic; cardiant; cardioglycoside; pRH3-1; pRH4-23; pRH4-31.
XX
OS Rattus rattus.
XX
FH Key Location/Qualifiers
FT CDS 196..6253
FT /*tag= a
XX
XX US5380836-A.
XX
XX 10-JAN-1995.
XX
XX 13-FEB-1989; 89US-0331330.
XX
XX 13-FEB-1989; 89US-0331330.
XX
XX 30-SEP-1991; 91US-0768107.
XX
XX (ARCH-) ARCH DEV CORP.
XX
XX Rogart RB;
XX
XX WPI; 1995-060381/08.
XX
XX P-PSDB; AAR67913.
XX
XX Purified DNA's encoding rat and human cardiac sodium channel
XX protein - useful for recombinant expression to produce sodium
XX channel proteins.
XX
XX Claim 8; Fig 1a-1n; 39pp; English.
XX
CC The cDNA is derived from 3 overlapping cDNA clones, designated
CC plasmid pRH3-1 (ATCC 67885), plasmid pRH4-23 (ATCC 67886) and
CC plasmid pRH4-31 (ATCC 67887). A virus/circular DNA plasmid vector
CC comprising the cDNA may be transformed or transfected into a
CC prokaryote/eukaryote host cell, and the resulting recombinant sodium
CC channel protein has various therapeutic, diagnostic and prognostic
CC uses. It may also be used to develop more effective antiarrhythmic,
CC cardiant and cardioglycoside drugs.

```

```

XX SQ Sequence 7555 BP; 1577 A; 2309 C; 2103 G; 1566 T; 0 other;

Query Match 75.8%; Score 18.2; DB 16; Length 7555;
Best Local Similarity 87.0%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 tcagccagttccaccgccgtgcac 23
   ||||| ||||| ||||| |||||
Db 539 tcagcccttcaccgccgtgcgc 561

RESULT 3
AAQ73770/C
ID AAQ73770 standard; cDNA; 57 BP.
XX
AC AAQ73770;
XX
DT 26-MAY-1995 (first entry)
XX
DE Hepatitis C virus type-specific sequence.
XX
KW HCV-4; HCV-5; HCV-6; NS4; NS5; non-structural protein; antigen;
KW antibody; vaccine; diagnosis; therapy; typing; immunoassay; ss.
XX
OS Hepatitis C virus.
XX
PN WO9425602-A.
XX
PD 10-NOV-1994.
XX
PF 05-MAY-1994; 94WO-GB00957.
XX
PR 05-MAY-1993; 93GB-0009237.
PR 07-JAN-1994; 94GB-0000263.
XX
XX (COMM-) COMMON SERVICES AGENCY.
XX (MURE-) MUREX DIAGNOSTICS LTD.
XX
XX Pike IH, Simmonds P, Yap PL;
XX WPI; 1994-358278/44.
XX
XX New polynucleotide(s) specific for hepatitis C virus types 4, 5
XX and 6 - and related antigenic peptide(s) and antibodies, useful
XX in vaccines, diagnosis, HCV typing and treatment
XX
XX Claim 11; Page 57; 70pp; English.
XX
XX Viral RNA from HCV-infected patients was subjected to PCR
XX amplification using primers corresponding to highly conserved
XX sequences of non-structural proteins NS5 and NS4. Products were
XX cloned into pUC18 and expressed in Escherichia coli. NS5 cDNA
XX sequences unique to HCV-4 and HCV-6, and NS4 sequences unique to
XX HCV-4, HCV-5 and the newly discovered HCV-6 are given in AAQ73767-75.
XX
XX SQ Sequence 57 BP; 14 A; 16 C; 13 G; 14 T; 0 other;

Query Match 73.3%; Score 17.6; DB 15; Length 57;
Best Local Similarity 83.3%; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 tcagccagttccaccgccgtgcac 24
   ||||| ||||| ||||| |||||
Db 50 TCAGCAAGTTGCAACCCCGTGCTCG 27

RESULT 4
ABL24072/C
ID ABL24072 standard; DNA; 4541 BP.
XX

```

AC ABL24072;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 23689.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 XX 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 XX WPI; 2001-656860/75.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1; SEQ ID NO 23689; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (AB57737-AB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 4541 BP; 1232 A; 1097 C; 1013 G; 1199 T; 0 other;
 SQ

Query Match 72.5%; Score 17.4; DB 23; Length 4541;
 Best Local Similarity 94.7%; Pred. No. 2.7e+02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 agccagttccaccctgtgc 21
 ||||| ||||| ||||| |||||
 Db 1256 AGCCAAATTCACCCCGTGC 1238

RESULT 5
 AAH27139/c
 ID AAH27139 standard; DNA; 230 BP.
 AC
 XX
 XX AAH27139;
 DT 08-AUG-2001 (first entry)
 DE Human beta-2 adrenergic receptor UTR region with RBP binding ability.
 XX
 KW Untranslated region; UTR; RNA binding protein; RBP; neurodegeneration;
 KW stroke; cardiovascular disease; hypertension; cancer; inflammation;
 KW metabolic disorder; obesity; diabetes; beta-2 adrenergic receptor; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200134624-A1.

XX 17-MAY-2001.
 XX
 PF 09-NOV-2000; 2000WO-US30888.
 XX
 PR 10-NOV-1999; 99US-0437458.
 XX
 PA (MESS-) MESSAGE PHARM INC.
 XX
 PI Giordano A, Xavier AK;
 XX
 XX WPI; 2001-335904/35.
 XX
 PT New nucleic acids that bind RNA-binding proteins or regulate mRNA
 PT function, useful for therapeutic gene regulation, such as in cases of
 PT neurodegeneration -
 XX
 XX Claim 1; Page 28; 33pp; English.
 XX
 CC Sequences AAH27132 - AAH27151 represent human gene untranslated regions
 CC where the corresponding mRNA fragment has RNA binding protein (RBP)
 CC binding activity. RBPs mediate the processing of pre-mRNA, the transport
 CC of mRNA from the nucleus to the cytoplasm, mRNA stabilisation, therefore
 CC translational efficiency, and the sequestration of some mRNAs. Therefore
 CC modification of post-transcriptional protein expression in eukaryotic
 CC cells may be carried out through the targeting of specific interactions of
 CC proteins that bind to RBPs. The gene fragments of the invention are used
 CC to identify their optimized sub-fragments, compounds that affect RNA/RBP
 CC interaction or mRNA functionality; or RBPs that interact with the
 CC compounds. Compounds identified using the gene fragments are potentially
 CC useful for therapeutic regulation of gene expression, such as in cases of
 CC neurodegeneration; stroke; cardiovascular disease; hypertension; cancer;
 CC inflammation; metabolic disorders (obesity and diabetes) and bacterial or
 CC viral infection. The present sequence is one of gene fragments of the
 CC invention, isolated from the human beta-2 adrenergic receptor gene.
 XX
 XX Sequence 230 BP; 42 A; 91 C; 70 G; 27 T; 0 other;
 SQ

Query Match 71.7%; Score 17.2; DB 22; Length 230;
 Best Local Similarity 86.4%; Pred. No. 2.7e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 cagccagttccaccctgtgc 23
 | ||||| ||||| ||||| |
 Db 42 CTGCCAGTTCCAGCCCGTGC 21

RESULT 6
 AAQ54681/c
 ID AAQ54681 standard; cDNA to mRNA; 657 BP.
 XX
 AC AAQ54681;
 XX
 XX 06-JUL-1994 (first entry)
 DT
 XX
 DE Rat proteosome RING12 gene.
 XX
 KW Component; rat; proteosome; vector; transformant; tumours;
 KW diagnosis; expression vector; ss.
 XX
 OS Rattus rattus.
 XX
 PH Key Location/Qualifiers
 FT CDS 1..657
 FT /*tag= a
 FT /product= rat_proteosome_RING12
 XX
 PN JP05317059-A.
 XX
 XX 03-DEC-1993.
 PD
 XX 22-MAY-1992; 92JP-0154184.
 PF

XX 22-MAY-1992; 92JP-0154184.
 XX (BIOM-) BIO MATERIAL KENYUSHO KK.
 XX WPI: 1994-011024/02.
 DR P-PSDB; AAR47473.
 XX A component of rat proteosome - for elucidation of mechanism of
 PT e.g. malignant tumour
 PT Claim 8; Page 5-6; 24pp; Japanese.
 XX Sequences (AA054678-81) show various components of a rat proteosome.
 CC The component is useful for the elucidation of various mechanisms
 CC of diseases such as malignant tumours and their diagnosis and
 CC treatment.
 XX Sequence 657 BP; 137 A; 184 C; 198 G; 138 T; 0 other;
 SQ

Query Match 71.7%; Score 17.2; DB 15; Length 657;
 Best Local Similarity 86.4%; Pred. No. 2.9e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 tcagccagttccaccctgtgca 22
 || ||||| ||||| ||||| |||||
 Db 275 TCCTCCAGTTCACCGGTGCA 254

RESULT 7
 AAS19511
 ID AAS19511 standard; DNA; 1785 BP.
 XX AC AAS19511;
 XX 26-MAR-2002 (first entry)
 XX Reference sequence for human MPL gene exons 9-10.
 DE Human; single nucleotide polymorphism; SNP; MPL; chromosome 1p34;
 XX myeloproliferative leukaemia virus oncogene; haplotyping; genotyping;
 KW congenital amegakaryocytic thrombocytopaenia; CAMT; ds.
 XX Homo sapiens.
 OS
 XX Key Location/Qualifiers
 FH 614..773
 FT exon /*tag= a
 FT /*number= 9
 FT intron 774..11027
 FT /*tag= b
 FT /*number= 9
 FT exon 1028..11124
 FT /*tag= c
 FT /*number= 10
 XX WO200179232-A2.
 XX 25-OCT-2001.
 XX 16-APR-2001; 2001WO-US12301.
 XX 14-APR-2000; 2000US-197839P.
 XX (GENA-) GENAISSANCE PHARM INC.
 XX Chew A, Choi JY, Koshy B, Stephens JC;
 PI WPI: 2002-055251/07.
 DR P-PSDB; AAU09495.
 XX Nucleotide polymorphisms in the human myeloproliferative leukemia virus

PT oncogene (MPL) gene, useful for studying the function of and expressing
 PT MPL protein for use in screening drugs for treating diseases related to
 PT MPL activity -
 XX Claim 27; Fig 3; 85pp; English.
 XX The present invention relates to novel single nucleotide polymorphisms
 CC (SNPs) in the human myeloproliferative leukaemia virus oncogene (MPL)
 CC gene located on chromosome 1p34, and methods for haplotyping and/or
 CC genotyping the MPL gene. The methods of the invention make use of
 CC allele-specific oligonucleotides (ASOs) as probes and primers and/or
 CC primer-extension oligonucleotides for detecting MPL gene polymorphisms.
 CC The polynucleotides and screened compounds are useful for the
 CC treatment of diseases associated with MPL activity, such as
 CC congenital amegakaryocytic thrombocytopaenia (CAMT). The present
 CC sequence represents a reference sequence for human MPL gene exons 9-10.
 CC Note: This sequence encodes for only part of the MPL protein shown
 CC in AAU09495.
 XX
 SQ Sequence 1785 BP; 317 A; 562 C; 500 G; 403 T; 3 other;
 SQ

Query Match 71.7%; Score 17.2; DB 24; Length 1785;
 Best Local Similarity 86.4%; Pred. No. 3e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 tcagccagttccaccctgtgca 22
 ||||| ||||| ||||| ||||| |||||
 Db 1672 tcagcgtctctaccctgtgca 1693

RESULT 8
 AAS91810/c
 ID AAS91810 standard; cDNA; 1972 BP.
 XX AC AAS91810;
 XX 13-FEB-2002 (first entry)
 XX DNA encoding novel human diagnostic protein #27614.
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX Homo sapiens.
 OS
 XX WO200175067-A2.
 XX 11-OCT-2001.
 XX 30-MAR-2001; 2001WO-US08631.
 XX 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.
 XX Drmanac RT, Liu C, Tang YT;
 PI WPI: 2001-639362/73.
 DR P-PSDB; ABG27623.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX Claim 1; SEQ ID No 27614; 103pp; English.
 PS The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1972 BP; 460 A; 523 C; 552 G; 437 T; 0 other;

Query Match 71.7%; Score 17.2; DB 23; Length 1972;
 Best Local Similarity 86.4%; Pred. No. 3.1e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 2 cagccaggttcaccccggtgca 23
 Db 705 CCGCCAGTCCACACCGTGCAC 684

RESULT 9
 AAT90830
 ID AAT90830 standard; cDNA; 1996 BP.

AC AAT90830;
 XX 15-JAN-1998 (first entry)
 DE cDNA encoding glucoamylase P.

XX Glucoamylase P; hormoconis resiniae; debranching activity; enzyme;
 KW dextrinase activity; Trichoderma reesei; starch granule; preservation;
 KW hydrolysis; gelatinised starch; glucose syrup; straight linear dextrin;
 KW saccharification; lignocellulosic material; sugar utilisation;
 KW feed additive; ss.
 XX Hormoconis resiniae.

XX Key Location/Qualifiers
 FH CDS 30..1880
 FT /*tag= a
 FT polyA_signal 2515..2520
 FT /*tag= b

XX US5655585-A.

XX 09-SEP-1997.
 XX 03-SEP-1992; 92US-0937789.

XX 12-AUG-1993; 93US-0104853.
 XX 03-SEP-1992; 92US-0937789.
 XX 07-FEB-1995; 95US-0385370.

XX (ALKO-) ALKO-YHTIOT OY.

XX Aho S, Fagerstroem R, Joutsjoki V, Korhola M, Nevalainen H;
 PI Torckeli H, Torckeli T, Vainio A;

XX WPI; 1997-456802/42.
 DR P-FSDB; AAW30155.

PT Hormoconis resiniae glucoamylase P gene construct - for transforming

PT Trichoderma to produce recombinant glucoamylase P
 XX
 PS Claim 2; Column 55-58; 6lpp; English.

XX AAT90830 and AAT90831 represent the cDNA and gene encoding the
 CC Hormoconis resiniae glucoamylase P, respectively. These sequences are used
 CC in the composition of the invention and are capable of being processed by
 CC a Trichoderma host cell. H. resiniae glucoamylase P has higher
 CC debranching and dextrinase activity than conventional
 CC glucoamylase-pullulanase mixtures. T. reesei secretes enzymes that are
 CC important for the degradation of complexes around and in starch granules.
 CC The recombinant glucoamylase P enzyme produced by the Trichoderma host
 CC cell is useful in applications requiring the hydrolysis of gelatinised
 CC starch or the presence of a debranching activity. These applications
 CC include starch analysis, the manufacture of glucose syrups, production of
 CC straight linear dextrin for use in food, medicines and cosmetics, and in
 CC the preparation of food fibers by the enzymatic treatment of seed husks
 CC or brans. The enzyme can also be used as an additive to laundry and dish
 CC washing detergents, in wood and textile industry applications such as
 CC the preparation of plywood adhesives. It can also be used in the
 CC saccharification of lignocellulosic materials, the preservation of
 CC protein-containing animal or vegetable fodder, the preparation of
 CC feedstuff from meat by-products and to improve utilisation of sugar
 CC present in such feedstuff. When used as a feed additive for food
 CC animals, the recombinant bacterial biomass may be added directly to the
 CC animal's fodder.
 XX
 SQ Sequence 1996 BP; 474 A; 574 C; 491 G; 457 T; 0 other;

Query Match 71.7%; Score 17.2; DB 18; Length 1996;
 Best Local Similarity 86.4%; Pred. No. 3.1e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 tcagccagttccaccccggtgca 22
 Db 1473 tcgcagctccaccccggtgca 1494

RESULT 10

AX61116/c

ID AAX61116 standard; DNA; 2300 BP.

XX AAX61116;

XX 27-JUL-1999 (first entry)

XX Human beta2-adrenergic receptor gene.

XX AlphaB-adrenergic receptor; human; cardiovascular disease;
 KW beta2 adrenergic receptor; genetic variation identification; hypertrophy;
 KW disease diagnosis; hypertension; prostatic disease; pulmonary disorder;
 KW asthma; peripheral vascular disorder; neuropsychic disorder;
 KW endocrine-metabolic disorder; ss.

XX Homo sapiens.

XX WO9244454-A1.

XX 20-MAY-1999.

XX 04-NOV-1998; 98WO-US23496.

XX 10-NOV-1997; 97US-0086232.

XX (REGC) UNIV CALIFORNIA.

XX Buescher R, Herrmann V, Insel PA;

XX WPI; 1999-327357/27.

XX Pairs of oligonucleotides for amplifying adrenergic receptor genes

PS Disclosure; Page 124-125; 126pp; English.

Query Match	71.7%;	Score 17.2;	DB 21;	Length 2305;					
Best Local Similarity	86.4%;	Pred. No. 3.1e+02;							
Matches 19;	Conservative 0;	Mismatches 3;	Indels 0;	Gaps 0;					
QY	2	cagcagcttcaccccgtagc	23						
Db	616	CTCCAGTTCACGCCGTGCTC	595						
RESULT	12								
AAA38784/c									
ID	AAA38784	standard;	DNA;	2340 BP.					
XX									
AC	AAA38784;								
XX									
DT	05-OCT-2000	(first entry)							
XX									
DE	Human beta2 adrenergic receptor	beta2AR gene.							
XX									
KW	Human;	adrenergic receptor; beta2 adrenergic receptor; beta2AR;							
KW	chromosome 5q31(12);	disease predisposition; asthma; hypertension;							
KW	congestive heart failure;	ischemic heart disease; arrhythmia;							
KW	obesity; diabetes;	vascular disease; premature labour; migraine;							
KW	anaphylaxis;	chronic obstructive pulmonary disease; ds.							
XX									
OS	Homo sapiens.								
XX									
Key	Location/Qualifiers								
FF	1487..2340								
CDS	/*tag= a								
FF	/*product= "beta2 adrenergic receptor"								
FT									

FT /note= "no stop codon given at 3' end of sequence"

FT /partial

FT 1487..1546

FT /tag= b

FT /label= 5'_leader_cistron

FT replace(1541,T)

FT /tag= c

FT 1588..2340

FT /tag= d

XX WC200031307-A1.

XX

XX 02-JUN-2000.

XX

XX 24-NOV-1999; 99WO-US27963.

XX

XX 25-NOV-1998; 98US-0109886.

XX

XX (UYCI-) UNIV CINCINNATI.

XX

XX Liggett SB;

XX

XX WPI; 2000-400107/34.

XX

XX Polymorphisms in the leader cistron (LC) of the beta 2-adrenergic

PT receptor (beta 2 AR), useful for predicting genetic disposition to a

PT disease modified by beta 2 AR expression e.g. congestive heart failure,

PT hypertension -

XX

PS Disclosure; Figure 1; 56pp; English.

XX

CC The present sequence is a fragment of the C allele of the human beta2

CC adrenergic receptor (beta2AR) gene, which is located on chromosome

CC 5q31 (12). The gene has two different alleles, and it has been shown that

CC the presence of two copies of the r allele leads to higher expression of

CC the gene. This is because the polymorphism is found in the 5' leader

CC sequence, which encodes a peptide which regulates expression of the

CC beta2AR gene. The polymorphism is thought to affect individuals'

CC responses to beta-agonists and beta-antagonists, and is likely to

CC influence their predisposition to asthma, hypertension,

CC congestive heart failure, ischemic heart disease, arrhythmia, obesity,

CC diabetes, vascular disease, premature labour, migraine, anaphylaxis and

CC chronic obstructive pulmonary disease (COPD). The gene can, therefore, be

CC used to predict the susceptibility of an individual to these diseases and

CC determine the best treatment.

XX

XX Sequence 2340 BP; 498 A; 627 C; 653 G; 562 T; 0 other;

Query Match 71.7%; Score 17.2; DB 21; Length 2340;

Best Local Similarity 86.4%; Pred. No. 3.1e+02;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 cagccagttccaccocgtgcac 23

| ||||| ||||| ||||| |||||

Db 1410 CTGCCAGTTCAGCCCGTGCTC 1389

RESULT 13

AAT90831

ID AAT90831 standard; cDNA; 2745 BP.

XX AC AAT90831;

XX DT 15-JAN-1998 (first entry)

XX Glucoamylase P gene.

XX Glucoamylase P; hormoconis resinae; debranching activity; enzyme;

KW dextrinase activity; Trichoderma reesei; starch granule; preservation;

KW hydrolysis; gelatinised starch; glucose syrup; straight linear dextrin;

KW saccharification; lignocellulosic material; sugar utilisation;

KW feed additive; ss.

XX Hormoconis resinae.

OS

XX Key Location/Qualifiers

PH 5'UTR 1..294

FT /tag= a

FT CAAT_signal 87..90

FT /tag= b

FT TATA_signal 225..229

FT /tag= c

FT exon 295..520

FT /tag= d

FT /number= 1

FT /note= "encodes residues 1 to 75"

FT intron 521..593

FT /tag= e

FT /number= 1

FT /note= "designated IVS (intervening sequence) 1"

FT exon 594..898

FT /tag= f

FT /number= 2

FT /note= "encodes residues 76 to 177"

FT intron 899..1074

FT /tag= g

FT /number= 2

FT /note= "designated IVS (intervening sequence) 2"

FT exon 1075..1583

FT /tag= h

FT /number= 3

FT /note= "encodes residues 178 to 347"

FT intron 1584..1636

FT /tag= i

FT /number= 3

FT /note= "designated IVS (intervening sequence) 3"

FT exon 1637..2447

FT /tag= j

FT /number= 4

FT /note= "encodes residues 348 to 616"

FT 3'UTR 2448..2745

FT /tag= k

FT polyA_signal 2515..2520

FT /tag= l

XX US5665585-A.

XX 09-SEP-1997.

XX

XX 03-SEP-1992; 92US-0937789.

XX

XX 12-AUG-1993; 93US-0104853.

PR 03-SEP-1992; 92US-0937789.

PR 07-FEB-1995; 95US-0385370.

XX

PA (ALKO-) ALKO-YHTIOT OY.

XX

PI Aho S, Fagerstroem R, Joutsjoki V, Korhola M, Nevalainen H;

PI Torckeli H, Torckeli T, Vainio A;

XX

DR WPI; 1997-456802/42.

DR P-PSDB; AAW30155.

XX

PT Hormoconis resinae glucoamylase P gene construct - for transforming

PT Trichoderma to produce recombinant glucoamylase P

XX

PS Claim 2; Column 60-66; 61pp; English.

XX

CC AAT90830 and AAT90831 represent the cDNA and gene encoding the

CC Hormoconis resinae glucoamylase P, respectively. These sequences are used

CC in the composition of the invention and are capable of being processed by

CC a Trichoderma host cell. H. resinae glucoamylase P has higher

CC debranching and dextrinase activity than conventional

CC glucoamylase-pullulanase mixtures. T. reesei secretes enzymes that are

CC important for the degradation of complexes around and in starch granules.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 14:01:55 ; Search time 6034.22 Seconds
(without alignments)
83.231 Million cell updates/sec

Title: US-09-899-718a-7
Perfect score: 24
Sequence: 1 tcagccagttccaccgccgtgcacg 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_hhg.*
- 3: gb_in.*
- 4: gb_ov.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vl.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vl.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

1	24	100.0	24	6	AX349069
2	24	100.0	3785	6	AX349063
3	19.2	80.0	177696	2	AC103298
c	4	18.8	78.3	33	AC028268
	5	18.8	78.3	33	AC080286
	6	18.8	78.3	33	AC072500
c	7	18.8	78.3	33	AC070455
	8	18.8	78.3	33	AC0591790
c	9	18.4	76.7	33	AC089614
c	10	18.4	76.7	33	AC101702
c	11	18.4	76.7	33	AC099583
c	12	18.2	75.8	33	AC074630
c	13	18.2	75.8	33	AC062755
	14	18.2	75.8	33	AC074924
	15	18.2	75.8	33	AC074924
	16	18.2	75.8	33	AC074924
	17	18.2	75.8	33	AC074924
	18	18.2	75.8	33	AC074924
	19	18.2	75.8	33	AC074924
c	20	18.2	75.8	33	AC074924
c	21	18.2	75.8	33	AC074924
c	22	18.2	75.8	33	AC074924
	23	18.2	75.8	33	AC074924
	24	18.2	75.8	33	AC074924
	25	18.2	75.8	33	AC074924
	26	18.2	75.8	33	AC074924
	27	18.2	75.8	33	AC074924
	28	18.2	75.8	33	AC074924
	29	18.2	75.8	33	AC074924
c	30	18.2	75.8	33	AC074924
c	31	18.2	75.8	33	AC074924
c	32	18.2	75.8	33	AC074924
c	33	18.2	75.8	33	AC074924
c	34	18.2	75.8	33	AC074924
c	35	18.2	75.8	33	AC074924
c	36	18.2	75.8	33	AC074924
c	37	18.2	75.8	33	AC074924
c	38	18.2	75.8	33	AC074924
c	39	18.2	75.8	33	AC074924
c	40	18.2	75.8	33	AC074924
c	41	18.2	75.8	33	AC074924
c	42	18.2	75.8	33	AC074924
c	43	18.2	75.8	33	AC074924
c	44	18.2	75.8	33	AC074924
c	45	18.2	75.8	33	AC074924

ALIGNMENTS

RESULT	1	AX349069	Sequence	7 from Patent WO0202785.	24 bp	DNA	linear	PAT 06-FEB-2002
LOCUS	AX349069	Sequence	7 from Patent WO0202785.	24 bp	DNA	linear	PAT 06-FEB-2002	
DEFINITION	AX349069	Sequence	7 from Patent WO0202785.	24 bp	DNA	linear	PAT 06-FEB-2002	
ACCESSION	AX349069	Sequence	7 from Patent WO0202785.	24 bp	DNA	linear	PAT 06-FEB-2002	
VERSION	AX349069.1	GI:18615104						
KEYWORDS		synthetic construct.						
SOURCE		synthetic construct						
ORGANISM		artificial sequence.						
REFERENCE		1 (sites)						
AUTHORS		Sprunck,S., Kluth,A., Becker,D., Luetticke,S. and Loerz,H.						
TITLE		Promoters of gene expression in plant caryopses						
JOURNAL		Patent: WO 0202785-A 7 10-JAN-2002;						
FEATURES		Aventis CropScience GmbH (DE)						
source		Location/Qualifiers						
		1..24						
		/organism="synthetic construct"						
		/db_xref="taxon:32630"						
		/note="Oligonucleotide"						
		4 a 11 c 5 g 4 t						

BASE COUNT 4 a 11 c 5 g 4 t
ORIGIN


```
* 38705 42885: contig of 4181 bp in length
* 42886 42985: gap of unknown length
* 42986 42986: contig of 3905 bp in length
* 46890 46890: gap of unknown length
* 46891 46891: contig of 4599 bp in length
* 51589 51589: gap of unknown length
* 51590 51590: contig of 4587 bp in length
* 51690 51690: gap of unknown length
* 56277 56277: contig of 4269 bp in length
* 56377 56377: gap of unknown length
* 60645 60645: contig of 4861 bp in length
* 60646 60646: gap of unknown length
* 60746 60746: contig of 4861 bp in length
* 65607 65607: gap of unknown length
* 65707 65707: contig of 5480 bp in length
* 71187 71187: gap of unknown length
* 71287 71287: contig of 3092 bp in length
* 74379 74379: gap of unknown length
* 74479 74479: contig of 3240 bp in length
* 77719 77719: gap of unknown length
* 77819 77819: contig of 1909 bp in length
* 79728 79728: gap of unknown length
* 79827 79827: contig of 3217 bp in length
* 83044 83044: gap of unknown length
* 83144 83144: contig of 3028 bp in length
* 83145 83145: gap of unknown length
* 86172 86172: gap of unknown length
* 86272 86272: contig of 3640 bp in length
* 89912 89912: gap of unknown length
* 90012 90012: contig of 1937 bp in length
* 90013 90013: gap of unknown length
* 91949 91949: contig of 3037 bp in length
* 91950 91950: gap of unknown length
* 92049 92049: contig of 2903 bp in length
* 95087 95087: gap of unknown length
* 95186 95186: contig of 2903 bp in length
* 95887 95887: gap of unknown length
* 98089 98089: contig of 1916 bp in length
* 98189 98189: gap of unknown length
* 98190 98190: contig of 2282 bp in length
* 100106 100106: gap of unknown length
* 100205 100205: contig of 2885 bp in length
* 102488 102488: gap of unknown length
* 102587 102587: contig of 2885 bp in length
* 102588 102588: gap of unknown length
* 105473 105473: contig of 1383 bp in length
* 105572 105572: gap of unknown length
* 106956 106956: contig of 1383 bp in length
* 107055 107055: gap of unknown length
* 107056 107056: contig of 1938 bp in length
* 108994 108994: gap of unknown length
* 109094 109094: contig of 2285 bp in length
* 111379 111379: gap of unknown length
* 111479 111479: contig of 2024 bp in length
* 113503 113503: gap of unknown length
* 113602 113602: contig of 2360 bp in length
* 113603 113603: gap of unknown length
* 115963 115963: contig of 2600 bp in length
* 116063 116063: gap of unknown length
* 118663 118663: contig of 3114 bp in length
* 118763 118763: gap of unknown length
* 121877 121877: contig of 2720 bp in length
* 121977 121977: gap of unknown length
* 124697 124697: contig of 1911 bp in length
* 124797 124797: gap of unknown length
* 126707 126707: contig of 2181 bp in length
* 126807 126807: gap of unknown length
* 126808 126808: contig of 2181 bp in length
* 128989 128989: gap of unknown length
* 131431 131431: contig of 2343 bp in length
* 131432 131432: gap of unknown length
* 131532 131532: contig of 1124 bp in length
* 132556 132556: gap of unknown length
* 132756 132756: contig of 1242 bp in length
* 133998 133998: gap of unknown length
* 134098 134098: contig of 1849 bp in length
* 135947 135947: gap of unknown length
* 136047 136047: contig of 1732 bp in length
* 137779 137779: gap of unknown length
* 137879 137879: contig of 1233 bp in length
* 139211 139211: gap of unknown length
* 139212 139212: contig of 1239 bp in length
* 140451 140451: gap of unknown length
* 140551 140551: contig of 2192 bp in length
* 142742: contig of 2192 bp in length
* 142743 142743: gap of unknown length
* 142843 142843: contig of 2046 bp in length
* 144889 144889: gap of unknown length
* 144989 144989: contig of 1173 bp in length
* 146162 146162: gap of unknown length
* 146262 146262: contig of 1932 bp in length
* 148194 148194: gap of unknown length
* 148294 148294: contig of 1276 bp in length
* 149569 149569: gap of unknown length
* 149670 149670: contig of 1309 bp in length
* 150978 150978: gap of unknown length
* 151079 151079: contig of 2012 bp in length
* 153091 153091: gap of unknown length
* 153191 153191: contig of 1903 bp in length
* 155094 155094: gap of unknown length
* 155194 155194: contig of 1127 bp in length
* 156321 156321: gap of unknown length
* 156421 156421: contig of 1525 bp in length
* 157945 157945: gap of unknown length
* 158045 158045: contig of 1417 bp in length
* 158462 158462: gap of unknown length
* 159563 159563: contig of 1013 bp in length
* 160576 160576: gap of unknown length
* 160676 160676: contig of 1490 bp in length
* 162166 162166: gap of unknown length
* 162266 162266: contig of 1378 bp in length
* 163644 163644: gap of unknown length
Query Match 80.0%; Score 19.2; DB 2; Length 177696;
Best Local Similarity 87.5%; Pred. No. 2.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 tcagccagttccaccctgtgacg 24
Db 91657 TCAGCCCTCTTCACCCCTGTGCACG 91680
RESULT 4
AC028268/C standard; DNA; HTG; 959 BP.
ID AC028268
XX AC AC028268;
XX SV AC028268.1
XX 05-APR-2000 (Rel. 63, Created)
DT 05-APR-2000 (Rel. 63, Last updated, Version 1)
XX Giardia intestinalis clone CI0266 strain WB-C6, LOW-PASS SEQUENCE SAMPLING.
XX HTG; HTGS_PHASE0.
XX Giardia intestinalis
OC Eukaryota; Diplomonadida; Hexamitidae; Giardia.
XX [1]
XX 1-959
XX Morrison H.G., McArthur A.G., Nixon J., Eakin N.O., Kim U., Crocker M.C.,
XX Hinkle G., Holder M.E., Sogin M.L.;
XX "Giardia: a model for ancient eukaryotic genome analysis";
XX Unpublished.
XX [2]
XX 1-959
XX Crocker M.K., Morrison H.G., McArthur A.G., Nixon J., Eakin N.O., Kim U.,
XX Hinkle G., Holder M.E., Sogin M.L.;
XX Submitted (01-APR-2000) to the EMBL/GenBank/DBJ databases.
XX Josephine Bay Paul Center for Comparative Molecular Biology and Evolution,
XX Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543-1015, USA
CC * NOTE: This record contains 1 individual
CC * sequencing reads that have not been assembled into
```


	XX	SQ	Sequence 990 BP; 228 A; 273 C; 282 G; 206 T; 1 other;
		Query Match	78.3%; Score 18.8; DB 33; Length 990;
		Best Local Similarity	90.9%; Pred. NO. 6.7e+02;
		Mismatches	20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
	QY	3 acccagttccacccccgtgcacg 24 	
	Dd	800 AGCCAGATCGACCCGTCACG 779 	
	RESULT 8		
	SME591790	323450 bp DNA linear BCT 16-AUG-2001	
	LOCUS	Sinorhizobium meliloti 1021 complete chromosome; segment 9/12.	
	DEFINITION	AL591790 AL591688	
	VERSION	AL591790.1 GI:15075230	
	KEYWORDS	.	
	SOURCE	Sinorhizobium meliloti.	
	ORGANISM	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Rhizobiaceae; Sinorhizobium. Capela,D., Barloy-Hubler,F., Gouzy,J., Bothe,G., Ampe,F., Batut,J., Boistard,P., Becker,A., Boutry,M., Cadieu,E., Dreano,S., Gloux,S., Godrie,T., Goftau.A., Kahn,D., Kiss,E., Lelaur,V., Masuy,D., Pohl,I., Portetelle,D., Puechier,A., Purnelle,B., Ramsperger,U., Renard,C., Thebaud,P., Vandenbol,M., Weidner,S. and Galibert.F. From the Cover: Analysis of the chromosome sequence of the legume symbiont Sinorhizobium meliloti strain 1021 Proceedings of the National Academy of Sciences of the United States of America. 98 (17), 9877-9882 (2001)	
	TITLE	JOURNAL	PUBMED
	AUTHORS	Gouzy,J.	REFERENCE
		Direct Submission	EU Consortium
		MELILO EU Consortium:	COMMENT
		Laboratoire de Biologie Molculaire des Relations Plantes-Microorganismes, UMR215-CNRS-INRA, BP27, F-31326 Castanet, France, Laboratoire de Genetique et Developpement UMR6061-CNRS, Faculte de Medecine, 2 avenue du Pr. Leon Bernard, F-33043 Rennes, France, CATC GmbH, Fritz-Arnold-str. 23, D-78467 Konstanz, Germany, Universitaet Bielefeld, Biologie IV (Genetik) Universitaetsstr 25, D-33615 Bielefeld, Germany, Unite de Biochimie physiologique, Universite Catholique de Louvain, Place Croix du Sud 2, Bte 20, B-1348 Louvain-la-Neuve, Belgium, Unites Microbiologie, Faculte des Sciences Agronomiques de Gembloux, Avenue Marechal Juin 6, B-5030 Gembloux, Belgium. E-mail:jermone.gouzy@univ-louvain.be http://sequence.toulouse.inra.fr/melioti.html . Location/Qualifiers 1..323450 /organism="Sinorhizobium meliloti" /strain="1021" /db_xref=taxon:382 155..1519 /gene=SMC01575 155..1519 /gene=SMC01575 /function="miscellaneous; hypothetical/partial homology"	
	FEATURES	source	gene
	CDS		
	FH	Key	Location/Qualifiers
	FT	1..990	/db_xref="taxon:5741"
	FT	/organism="Giardia intestinalis"	
	FT	/strain="WB-C6"	
	FT	/clone="EJ6686"	
	FX	Source	
		Query Match	78.3%; Score 18.8; DB 33; Length 972;
		Best Local Similarity	90.9%; Pred. No. 6.7e+02;
		Mismatches	20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
	QY	3 acccagtccaacccttcgccacg 24 	
	Dd	460 AGCCAGATCGACCCGTGCACG 481 	
	RESULT 7		
	ID AC070455/C	standard; DNA; HTG; 990 BP.	
	XX AC070455;		
	SV AC070455.1		
	DT 13-JUN-2000 (Rel. 64, Created)		
	DD 13-JUN-2000 (Rel. 64, Last updated, Version 1)		
	DE Giardia intestinalis clone EJ6686 strain WB-C6, LOW-PASS SEQUENCE SAMPLING.		
	DX HTG; HTGS_PHASE0.		
	OS Giardia intestinalis		
	RN Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.		
	RP [1]		
	RR Morrison H.G., McArthur A.G., Nixon J., Eakin N.Q., Kim U., Crocker M.C.,		
	RT Hinkle G., Holder M.E., Sogin M.L.;		
	RL "Giardia: a model for ancient eukaryotic genome analysis";		
	RV Unpublished.		
	[2]		
	Eakin N.Q., Morrison H.G., McArthur A.G., Nixon J., Kim U., Crocker M.C.,		
	Hinkle G., Holder M.E., Sogin M.L.;		
	: Submitted (06-JUN-2000) to the EMBL/GenBank/DBJ databases.		
	Josephine Bay Paul Center for Comparative Molecular Biology and Evolution,		
	Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543-1015, USA		
	* NOTE: This record contains 1 individual		
	** sequencing reads that have not been assembled into		
	** contigs. Runs of N are used to separate the reads		
	** and the order in which they appear is completely		
	** arbitrary. Low-pass sequence sampling is useful for		
	** identifying clones that may be gene-rich and allows		
	** overlap relationships among clones to be deduced.		
	** However, it should not be assumed that this clone		
	** will be sequenced to completion. In the event that		
	** the record is updated, the accession number will		
	** be preserved.		
	1	990: contig of 990 bp in length.	
	XH Location/Qualifiers		
	FF FT	1..990	/db_xref="taxon:5741"
	FF FT	/organism="Giardia intestinalis"	
	FF FT	/strain="WB-C6"	
	FF FT	/clone="EJ6686"	

* 3793 3892: gap of 100 bp
 * 3893 4571: contig of 679 bp in length
 * 4572 4671: gap of 100 bp
 * 4672 5366: contig of 695 bp in length
 * 5367 5466: gap of 100 bp
 * 5467 6158: contig of 692 bp in length
 * 6159 6258: gap of 100 bp
 * 6259 6962: contig of 704 bp in length
 * 6963 7062: gap of 100 bp
 * 7063 7750: contig of 688 bp in length
 * 7751 7850: gap of 100 bp
 * 7851 8552: contig of 702 bp in length
 * 8553 8652: gap of 100 bp
 * 8653 9382: contig of 730 bp in length
 * 9383 9482: gap of 100 bp
 * 9483 10192: contig of 710 bp in length
 * 10193 10292: gap of 100 bp
 * 10293 10966: contig of 674 bp in length
 * 10967 11066: gap of 100 bp
 * 11067 11762: contig of 696 bp in length
 * 11763 11862: gap of 100 bp
 * 11863 12554: contig of 692 bp in length
 * 12555 12654: gap of 100 bp
 * 12655 13354: contig of 700 bp in length
 * 13355 13454: gap of 100 bp
 * 13455 14149: contig of 695 bp in length
 * 14150 14249: gap of 100 bp
 * 14250 14972: contig of 723 bp in length
 * 14973 15072: gap of 100 bp
 * 15073 15740: contig of 668 bp in length
 * 15741 15840: gap of 100 bp
 * 15841 16550: contig of 710 bp in length
 * 16551 16650: gap of 100 bp
 * 16651 17333: contig of 683 bp in length
 * 17334 17433: gap of 100 bp
 * 17434 18134: contig of 701 bp in length
 * 18135 18234: gap of 100 bp
 * 18235 18924: contig of 690 bp in length
 * 18925 19024: gap of 100 bp
 * 19025 19703: contig of 679 bp in length
 * 19704 19803: gap of 100 bp
 * 19804 20518: contig of 715 bp in length
 * 20519 20618: gap of 100 bp
 * 20619 21332: contig of 714 bp in length
 * 21333 21432: gap of 100 bp
 * 21433 22134: contig of 702 bp in length
 * 22135 22234: gap of 100 bp
 * 22235 22917: contig of 683 bp in length
 * 22918 23017: gap of 100 bp
 * 23018 23708: contig of 691 bp in length
 * 23709 23808: gap of 100 bp
 * 23809 24495: contig of 687 bp in length
 * 24496 24595: gap of 100 bp
 * 24596 25296: contig of 701 bp in length
 * 25297 25396: gap of 100 bp
 * 25397 26102: contig of 706 bp in length
 * 26103 26202: gap of 100 bp
 * 26203 26926: contig of 724 bp in length
 * 26927 27026: gap of 100 bp
 * 27027 27694: contig of 668 bp in length
 * 27695 27794: gap of 100 bp
 * 27795 28517: contig of 723 bp in length
 * 28518 28617: gap of 100 bp
 * 28618 29334: contig of 717 bp in length
 * 29335 29434: gap of 100 bp
 * 29435 30116: contig of 682 bp in length
 * 30117 30216: gap of 100 bp
 * 30217 30903: contig of 687 bp in length
 * 30904 31003: gap of 100 bp
 * 31004 31693: contig of 690 bp in length
 * 31694 31793: gap of 100 bp
 * 31794 32483: contig of 690 bp in length
 * 32484 32583: gap of 100 bp

* 32584 33258: contig of 675 bp in length
 * 33259 33358: gap of 100 bp
 * 33359 34077: contig of 719 bp in length
 * 34078 34177: gap of 100 bp
 * 34178 34887: contig of 710 bp in length
 * 34888 34987: gap of 100 bp
 * 34988 35656: contig of 669 bp in length
 * 35657 35756: gap of 100 bp
 * 35757 36444: contig of 688 bp in length
 * 36445 36544: gap of 100 bp
 * 36545 37262: contig of 718 bp in length
 * 37263 37362: gap of 100 bp
 * 37363 38053: contig of 691 bp in length
 * 38054 38153: gap of 100 bp
 * 38154 38857: contig of 704 bp in length
 * 38858 38957: gap of 100 bp
 * 38958 39666: contig of 709 bp in length
 * 39667 39766: gap of 100 bp
 * 39767 40463: contig of 697 bp in length
 * 40464 40563: gap of 100 bp
 * 40564 41266: contig of 703 bp in length
 * 41267 41366: gap of 100 bp
 * 41367 42090: contig of 724 bp in length
 * 42091 42190: gap of 100 bp
 * 42191 42907: contig of 717 bp in length
 * 42908 43007: gap of 100 bp
 * 43008 43708: contig of 701 bp in length
 * 43709 43808: gap of 100 bp
 * 43809 44503: contig of 695 bp in length
 * 44504 44603: gap of 100 bp
 * 44604 45297: contig of 694 bp in length
 * 45298 45397: gap of 100 bp
 * 45398 46100: contig of 703 bp in length
 * 46101 46200: gap of 100 bp
 * 46201 46892: contig of 692 bp in length
 * 46893 46992: gap of 100 bp
 * 46993 47682: contig of 690 bp in length
 * 47683 47782: gap of 100 bp
 * 47783 48493: contig of 711 bp in length
 * 48494 48593: gap of 100 bp
 * 48594 49276: contig of 683 bp in length
 * 49277 49376: gap of 100 bp
 * 49377 50096: contig of 720 bp in length
 * 50097 50196: gap of 100 bp
 * 50197 50903: contig of 707 bp in length
 * 50904 51003: gap of 100 bp
 * 51004 51660: contig of 657 bp in length
 * 51661 51760: gap of 100 bp
 * 51761 52454: contig of 694 bp in length
 * 52455 52554: gap of 100 bp
 * 52555 53255: contig of 701 bp in length
 * 53256 53355: gap of 100 bp
 * 53356 54030: contig of 675 bp in length
 * 54031 54130: gap of 100 bp
 * 54131 54843: contig of 713 bp in length
 * 54844 54943: gap of 100 bp

Query Match 76.7%; Score 18.4; DB 2; Length 113288;
 Best Local Similarity 95.0%; Pred. No. 6.3e+02;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 ccagttccaccctgtgcacg 24
 |||||
 Db 99957 CCAGTTCACCCCGTGCACG 99938

RESULT 11

AC099583
 LOCUS AC099583 233684 bp DNA linear HTG 16-NOV-2001
 DEFINITION Mus musculus clone RP23-189G16, WORKING DRAFT SEQUENCE, 25
 unorderd pieces.
 ACCESSION AC099583
 VERSION AC099583.1 GI:16946032

KEYWORDS
SOURCEHTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
house mouse.

ORGANISM

Mus musculus
Mus musculus

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS

1 (bases 1 to 233684)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.

JOURNAL

Unpublished
Mus musculus, clone RP23-189G16

REFERENCE

2 (bases 1 to 233684)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigglio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE

Direct Submission

JOURNAL

Submitted (16-NOV-2001) Whitehead Institute/MIT Center for Genome

COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L16855

Center clone name: 189_G16

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 225791 bases at least Q40

Consensus quality: 229156 bases at least Q30

Consensus quality: 230361 bases at least Q20

Insert size: 240000; agarose-fp

Quality coverage: 11.5 in Q20 bases; agarose-fp

Quality coverage: 11.9 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently

* consists of 25 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence.

* as soon as it is available and the accession number will

* be preserved.

* 1 458: contig of 458 bp in length

* 459 558: gap of 100 bp

* 559 2100: contig of 1542 bp in length

* 2101 2200: gap of 100 bp

* 2201 3332: contig of 1132 bp in length

* 3333 3432: gap of 100 bp

* 3433 4952: contig of 1560 bp in length

* 4993 5092: gap of 100 bp
 * 5093 6602: contig of 1510 bp in length
 * 6603 6702: gap of 100 bp
 * 6703 10246: contig of 3544 bp in length
 * 10247 10346: gap of 100 bp
 * 10347 12129: contig of 1783 bp in length
 * 12130 12229: gap of 100 bp
 * 12230 14454: contig of 2225 bp in length
 * 14455 14554: gap of 100 bp
 * 14555 16558: contig of 2004 bp in length
 * 16559 16658: gap of 100 bp
 * 16659 19755: contig of 3097 bp in length
 * 19756 19855: gap of 100 bp
 * 19856 23363: contig of 3508 bp in length
 * 23364 23463: gap of 100 bp
 * 23464 27548: contig of 4085 bp in length
 * 27549 27648: gap of 100 bp
 * 27649 31362: contig of 3714 bp in length
 * 31363 31462: gap of 100 bp
 * 31463 36095: contig of 4637 bp in length
 * 36100 36199: gap of 100 bp
 * 36200 43798: contig of 7599 bp in length
 * 43799 43898: gap of 100 bp
 * 43899 51910: contig of 8012 bp in length
 * 51911 52010: gap of 100 bp
 * 52011 70202: contig of 18192 bp in length
 * 70203 70302: gap of 100 bp
 * 70303 102493: contig of 32191 bp in length
 * 102494 102593: gap of 100 bp
 * 102594 116654: contig of 14061 bp in length
 * 116655 116754: gap of 100 bp
 * 116755 129262: contig of 12508 bp in length
 * 129263 129362: gap of 100 bp
 * 129363 150868: contig of 21506 bp in length
 * 150869 150968: gap of 100 bp
 * 150969 176203: contig of 25235 bp in length
 * 176204 176303: gap of 100 bp
 * 176304 197144: contig of 20841 bp in length
 * 197145 197244: gap of 100 bp
 * 197245 221039: contig of 23795 bp in length
 * 221040 221339: gap of 100 bp
 * 221140 233684: contig of 12545 bp in length.

FEATURES

source
 1..233684
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="RP23-189G16"
 /clone_lib="RPC1-23 Female Mouse BAC"
 misc_feature
 1..458
 /note="assembly_fragment"
 vector_side:left
 clone_end:SP6
 misc_feature
 559..2100
 /note="assembly_fragment"
 misc_feature
 2201..3332
 /note="assembly_fragment"
 misc_feature
 3433..4992
 /note="assembly_fragment"
 misc_feature
 5093..6602
 /note="assembly_fragment"
 misc_feature
 6703..10246
 /note="assembly_fragment"
 misc_feature
 10347..12129
 /note="assembly_fragment"
 misc_feature
 12230..14454
 /note="assembly_fragment"
 misc_feature
 14555..16558
 /note="assembly_fragment"
 misc_feature
 16659..19755
 /note="assembly_fragment"
 misc_feature
 19856..23363
 /note="assembly_fragment"
 misc_feature
 23464..27548

RL	Josephine Bay Paul Center for Comparative Molecular Biology and Evolution
RM	Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543-1015, USA
XX	
CC	* NOTE: This record contains 1 individual
CC	* sequencing reads that have not been assembled into
CC	* contigs. Runs of N are used to separate the reads
CC	* and the order in which they appear is completely
CC	* arbitrary. Low-pass sequence sampling is useful for
CC	* identifying clones that may be gene-rich and allows
CC	* overlap relationships among clones to be deduced.
CC	* However, it should not be assumed that this clone
CC	* will be sequenced to completion. In the event that
CC	* the record is updated, the accession number will
CC	* be preserved.
CC	* 1 667: contig of 667 bp in length.
XX	
FF	Key Location/Qualifiers
FT	source 1. 667
FT	/db_xref="taxon:5741"
FT	/organism="Giardia intestinalis"
FT	/strain="WB-C6"
FT	/clone="AJ2486"
XX	
SQ	sequence 667 BP; 159 A; 143 C; 183 G; 176 T; 6 other;
Query Match 75.88; Score 18.2; DB 33; Length 667;	
Best Local Similarity 87.04; Pred. No. 1.3e+03;	
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;	
QY	1 tcagccaggttcaccccgctgcac 23
Db	661 TCAGCCAGTGCCTCCGCTGCAC 639
RESULT	13
HSAS32832	
LOCUS	
DEFINITION	HSA332832 Homo sapiens genomic sequence surrounding NotI site, clone
ACCESSION	HSJ-DG20RS.
VERSION	AJ332832
KEYWORDS	AJ332832.1 GI:15877250
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
	1 (bases 1 to 727)
	Kutsenko,A.S., Gizatullin,R.Z., Al-Amin,A.N., Wang,F.,
	Podowski,R.M., Matushkin,Y.G., Kvasha,S.M., Gyanchandani,A.,
	Muravenko,O.V., Protopopov,A.I., Kashuba,V.I., Kisselev,L.L.,
	Wasserman,W., Wählestedt,C. and Zabarovsky,E.R.
TITLE	Analysis of NotI flanking sequences: a new tool for gene discovery
	and verification of the human genome
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 727)
AUTHORS	Zabarovsky,E.R.
TITLE	Direct Submission
JOURNAL	Submitted (16-MAY-2001) Microbiology and Tumorbiology Centre,
	Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77,
	Sweden
FEATURES	Location/Qualifiers
source	1..727
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="HSJ-DG20RS"
BASE COUNT	166 a 239 c 217 g 104 t 1 others
ORIGIN	
Query Match 75.88; Score 18.2; DB 9; Length 727;	
Best Local Similarity 87.04; Pred. No. 1.3e+03;	

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 cagccaggttcaccccgtagc 24
||||| ||||| |||||
Db 674 CAGCCAGGTCCAGCGGTGCAC 696

RESULT 14
AC062755 standard; DNA; HTG; 904 BP.
XX AC062755;
XX AC062755.1
XX 26-APR-2000 (Rel. 63, Created)
DT 26-APR-2000 (Rel. 63, Last updated, Version 1)
XX Giardia intestinalis clone NJ4136 strain WB-C6, LOW-PASS SEQUENCE SAMPLING.
XX HTG; HTGS_PHASE0.
XX Giardia intestinalis
OS Eukaryota; Diplomonadida; Hexamitidae; Giardia.
XX [1]
RN 1-904
RP Morrison H.G., McArthur A.G., Nixon J., Eakin N.Q., Kim U., Crocker M.C.,
RA Hinkle G., Holder M.E., Sogin M.L.;
RT "Giardia: a model for ancient eukaryotic genome analysis";
RL Unpublished.
XX [2]
RN 1-904
RP Nixon J., Morrison H.G., McArthur A.G., Eakin N.Q., Kim U., Crocker M.C.,
RA Hinkle G., Holder M.E., Sogin M.L.;
RT Submitted (21-APR-2000) to the EMBL/GenBank/DBJ databases.
RL Josephine Bay Paul Center for Comparative Molecular Biology and Evolution,
RL Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543-1015, USA
XX

CC * NOTE: This record contains 1 individual
CC * sequencing reads that have not been assembled into
CC * contigs. Runs of N are used to separate the reads
CC * and the order in which they appear is completely
CC * arbitrary. Low-pass sequence sampling is useful for
CC * identifying clones that may be gene-rich and allows
CC * overlap relationships among clones to be deduced.
CC * However, it should not be assumed that this clone
CC * will be sequenced to completion. In the event that
CC * the record is updated, the accession number will
CC * be preserved.
XX 1 904: contig of 904 bp in length.
XX FH Location/Qualifiers
FH source 1. .904
FT /db.xref="taxon:5741"
FT /organism="Giardia intestinalis"
FT /strain="WB-C6"
FT /clone="NJ4136"
XX

XX Sequence 904 BP; 233 A; 237 C; 221 G; 213 T; 0 other;
Query Match 75.8%; Score 18.2; DB 33; Length 904;
Best Local Similarity 87.0%; Pred. No. 1.2e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 tcagccaggttcaccccgtagc 23
||||| ||||| |||||
Db 539 TCAGCCAGTCCCGCGGTGCAC 561

RESULT 15

AC074924 standard; DNA; HTG; 1102 BP.
XX AC074924;
XX AC074924.1
XX 02-AUG-2000 (Rel. 64, Created)
DT 02-AUG-2000 (Rel. 64, Last updated, Version 1)
XX Giardia intestinalis clone AJ2681 strain WB-C6, LOW-PASS SEQUENCE SAMPLING.
XX HTG; HTGS_PHASE0.
XX Giardia intestinalis
OS Eukaryota; Diplomonadida; Hexamitidae; Giardia.
XX [1]
RN 1-1102
RP Morrison H.G., McArthur A.G., Nixon J., Eakin N.Q., Kim U., Crocker M.C.,
RA Hinkle G., Holder M.E., Sogin M.L.;
RT "Giardia: a model for ancient eukaryotic genome analysis";
RL Unpublished.
XX [2]
RN 1-1102
RP McArthur A.G., Morrison H.G., Nixon J., Eakin N.Q., Kim U., Crocker M.C.,
RA Hinkle G., Holder M.E., Sogin M.L.;
RT Submitted (31-JUL-2000) to the EMBL/GenBank/DBJ databases.
RL Josephine Bay Paul Center for Comparative Molecular Biology and Evolution,
RL Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543-1015, USA
XX

CC * NOTE: This record contains 1 individual
CC * sequencing reads that have not been assembled into
CC * contigs. Runs of N are used to separate the reads
CC * and the order in which they appear is completely
CC * arbitrary. Low-pass sequence sampling is useful for
CC * identifying clones that may be gene-rich and allows
CC * overlap relationships among clones to be deduced.
CC * However, it should not be assumed that this clone
CC * will be sequenced to completion. In the event that
CC * the record is updated, the accession number will
CC * be preserved.
XX 1 1102: contig of 1102 bp in length.
XX FH Location/Qualifiers
FH source 1. .1102
FT /db.xref="taxon:5741"
FT /organism="Giardia intestinalis"
FT /strain="WB-C6"
FT /clone="AJ2681"
XX

XX Sequence 1102 BP; 294 A; 283 C; 258 G; 265 T; 2 other;
Query Match 75.8%; Score 18.2; DB 33; Length 1102;
Best Local Similarity 87.0%; Pred. No. 1.2e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 tcagccaggttcaccccgtagc 23
||||| ||||| |||||
Db 546 TCAGCCAGTCCCGCGGTGCAC 568

Search completed: July 31, 2002, 14:02:11
Job time: 17573 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 12:23:47 ; Search time 165.21 Seconds
(without alignments)
17.842 Million cell updates/sec

Title: US-09-899-718A-6
Perfect score: 12
Sequence: 1 caggagcctcga 12

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA: *
1: /cgn2.6/prodata/2/ina/5A_COMB.seq: *
2: /cgn2.6/prodata/2/ina/5B_COMB.seq: *
3: /cgn2.6/prodata/2/ina/6A_COMB.seq: *
4: /cgn2.6/prodata/2/ina/6B_COMB.seq: *
5: /cgn2.6/prodata/2/ina/PCTUS_COMB.seq: *
6: /cgn2.6/prodata/2/ina/Backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	12	100.0	475	US-09-247-155-147	Sequence 147, App
c 2	12	100.0	1341	US-09-199-637A-266	Sequence 266, App
c 3	12	100.0	1722	US-09-221-275-3	Sequence 3, Appli
c 4	12	100.0	2001	US-09-341-587-2	Sequence 2, Appli
c 5	12	100.0	2671	US-09-221-275-2	Sequence 2, Appli
c 6	12	100.0	2678	US-09-221-275-1	Sequence 1, Appli
c 7	12	100.0	3116	US-08-449-645A-14	Sequence 14, Appl
c 8	12	100.0	3116	US-08-702-367A-14	Sequence 14, Appl
c 9	12	100.0	3116	PCT-US95-04681-14	Sequence 14, Appl
c 10	12	100.0	3348	US-08-222-616-34	Sequence 34, Appl
c 11	12	100.0	3348	PCT-US95-04228-34	Sequence 34, Appl
c 12	12	100.0	3955	US-09-428-711A-13	Sequence 13, Appl
c 13	12	100.0	4680	US-08-254-358-1	Sequence 1, Appli
c 14	12	100.0	4680	US-08-475-391-1	Sequence 1, Appli
c 15	12	100.0	4680	US-08-709-609-1	Sequence 1, Appli
c 16	12	100.0	4680	PCT-US95-07178-1	Sequence 1, Appli
c 17	12	100.0	5802	US-09-341-587-4	Sequence 4, Appli
c 18	12	100.0	28720	US-09-341-587-7	Sequence 7, Appli
c 19	12	100.0	42235	US-09-199-637A-1	Sequence 1, Appli
c 20	11	91.7	30	US-08-836-504A-10	Sequence 10, Appl
c 21	11	91.7	97	US-08-370-975B-9	Sequence 9, Appli
c 22	11	91.7	334	US-08-358-171-22	Sequence 22, Appl
c 23	11	91.7	334	US-09-090-947-22	Sequence 22, Appl
c 24	11	91.7	340	US-08-594-031-83	Sequence 83, Appl
c 25	11	91.7	412	US-08-047-041A-1	Sequence 1, Appli
c 26	11	91.7	412	US-08-795-006A-21	Sequence 21, Appl
c 27	11	91.7	412	US-09-184-073-21	Sequence 21, Appl

c 28	11	91.7	492	2	US-08-892-880-12	Sequence 12, Appl
c 29	11	91.7	528	1	US-08-047-041A-14	Sequence 14, Appl
c 30	11	91.7	534	1	US-08-370-975B-13	Sequence 13, Appl
c 31	11	91.7	624	1	US-07-661-610C-14	Sequence 14, Appl
c 32	11	91.7	642	1	US-09-609-324A-9	Sequence 9, Appli
c 33	11	91.7	642	2	US-08-920-440B-9	Sequence 9, Appli
c 34	11	91.7	642	4	US-09-173-492-9	Sequence 9, Appli
c 35	11	91.7	642	4	US-09-173-133-9	Sequence 9, Appli
c 36	11	91.7	1025	4	US-09-269-617-13	Sequence 13, Appl
c 37	11	91.7	1329	1	US-08-378-630A-8	Sequence 8, Appli
c 38	11	91.7	1454	4	US-09-372-422A-19	Sequence 19, Appl
c 39	11	91.7	1560	1	US-08-356-180-1	Sequence 1, Appli
c 40	11	91.7	2313	2	US-08-892-880-1	Sequence 1, Appli
c 41	11	91.7	2419	2	US-08-765-662-13	Sequence 13, Appl
c 42	11	91.7	2419	5	PCT-US95-08745-13	Sequence 13, Appl
c 43	11	91.7	2542	3	US-08-941-445A-6	Sequence 6, Appli
c 44	11	91.7	3014	2	US-08-808-982-1	Sequence 1, Appli
c 45	11	91.7	3014	4	US-09-306-902A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-247-155-147/c
; Sequence 147, Application US/09247155A
; Patent No. 6312922
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert, Aymeric
; TITLE OF INVENTION: Complementary DNAs
; FILE REFERENCE: GENSET.021A
; CURRENT APPLICATION NUMBER: US/09/247,155A
; CURRENT FILING DATE: 1998-02-09
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-02-09
; EARLIER APPLICATION NUMBER: 60/081,563
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-08-10
; EARLIER APPLICATION NUMBER: 60/099,273
; EARLIER FILING DATE: 1998-10-04
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: Patent.pm
; SEQ ID NO 147
; LENGTH: 475
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 146..457
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 146..292
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 5.1999980926514
; OTHER INFORMATION: seq CFCLYPIPLCTS/HP
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 442..447
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 465..475
US-09-247-155-147

Query Match 100.0%; Score 12; DB 4; Length 475;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
|||||||

Db 223 CAGGAGCCTCGA 212

RESULT 2

US-09-199-637A-266/c
; Sequence 266, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard M.
; APPLICANT: Rahme, Laurence G.
; APPLICANT: Mahajan-Miklos, Shalina
; APPLICANT: Tan, Man-Wah
; APPLICANT: Cao, Hui
; APPLICANT: Drenkard, Eliana
; APPLICANT: Tsongalis, John
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; FILE REFERENCE: 00786/361002
; CURRENT APPLICATION NUMBER: US/09/199,637A
; PRIOR FILING DATE: 1998-11-25
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 266
; LENGTH: 1341
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-266

Query Match 100.0%; Score 12; DB 4; Length 1341;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggagcctcga 12
|||||

Db 669 CAGGAGCCTCGA 658

RESULT 3

US-09-221-275-3
; Sequence 3, Application US/09221275
; Patent No. 6329332
; GENERAL INFORMATION:
; APPLICANT: Borneman, William S.
; APPLICANT: Bodie, Elizabeth A.
; APPLICANT: Graycar, Thomas P.
; APPLICANT: Winetzk, Deborah S.
; TITLE OF INVENTION: Pleurotus Phenol Oxidizing Enzymes
; FILE REFERENCE: GC557
; CURRENT APPLICATION NUMBER: US/09/221,275
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1722
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: cDNA
US-09-221-275-3

Query Match 100.0%; Score 12; DB 4; Length 1722;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggagcctcga 12
|||||

Db 1230 caggagcctcga 1241

RESULT 4

US-09-341-587-2/c
; Sequence 2, Application US/09341587
; Patent No. 6346606
; GENERAL INFORMATION:
; APPLICANT: Mollenhauer, Jan
; TITLE OF INVENTION: Protein Containing an SRCR Domain
; FILE REFERENCE: 4121-108
; CURRENT APPLICATION NUMBER: US/09/341,587
; CURRENT FILING DATE: 1999-08-31
; EARLIER APPLICATION NUMBER: PCT/DE98/00096
; EARLIER FILING DATE: 1998-01-09
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2001
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-341-587-2

Query Match 100.0%; Score 12; DB 4; Length 2001;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggagcctcga 12
|||||

Db 248 CAGGAGCCTCGA 237

RESULT 5

US-09-221-275-2
; Sequence 2, Application US/09221275
; Patent No. 6329332
; GENERAL INFORMATION:
; APPLICANT: Borneman, William S.
; APPLICANT: Bodie, Elizabeth A.
; APPLICANT: Graycar, Thomas P.
; APPLICANT: Winetzk, Deborah S.
; TITLE OF INVENTION: Pleurotus Phenol Oxidizing Enzymes
; FILE REFERENCE: GC557
; CURRENT APPLICATION NUMBER: US/09/221,275
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 2671
; TYPE: DNA
; ORGANISM: Pleurotus ostreatus
US-09-221-275-2

Query Match 100.0%; Score 12; DB 4; Length 2671;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggagcctcga 12
|||||

Db 1976 caggagcctcga 1987

RESULT 6

US-09-221-275-1
; Sequence 1, Application US/09221275
; Patent No. 6329332
; GENERAL INFORMATION:
; APPLICANT: Borneman, William S.
; APPLICANT: Bodie, Elizabeth A.
; APPLICANT: Graycar, Thomas P.
; APPLICANT: Winetzk, Deborah S.
; TITLE OF INVENTION: Pleurotus Phenol Oxidizing Enzymes
; FILE REFERENCE: GC557
US-09-221-275-1

•

; CURRENT APPLICATION NUMBER: US/09/221,275
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2678
; TYPE: DNA
; ORGANISM: Pleurotus ostreatus
US-09-221-275-1

Query Match 100.0%; Score 12; DB 4; Length 2678;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
|||||

Db 1976 caggagcctcga 1987

RESULT 7
US-08-449-645A-14/C
; Sequence 14, Application US/08449645A
; Patent No. 5981245
; GENERAL INFORMATION:
; APPLICANT: Fox, Gary M.
; TITLE OF INVENTION: EPH-Like Receptor Protein Tyrosine
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Patent Operations/RBW
; STREET: 1840 Behavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/449,645A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-287
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3116 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 34..2994
US-08-449-645A-14

Query Match 100.0%; Score 12; DB 2; Length 3116;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
|||||

Db 731 CAGGAGCCTCGA 720

RESULT 8
US-08-702-367A-14/C
; Sequence 14, Application US/08702367A

; Patent No. 5981246
; GENERAL INFORMATION:
; APPLICANT: Fox, Gary M.
; TITLE OF INVENTION: EPH-Like Receptor Protein Tyrosine
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Patent Operations/RBW
; STREET: 1840 Behavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/702,367A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-287
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3116 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 34..2994
US-08-702-367A-14

Query Match 100.0%; Score 12; DB 2; Length 3116;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
|||||

Db 731 CAGGAGCCTCGA 720

RESULT 9
PCT-US95-04681-14/c
; Sequence 14, Application PC/TUS9504681
; GENERAL INFORMATION:
; APPLICANT: Fox, Gary M.
; TITLE OF INVENTION: EPH-Like Receptor Protein Tyrosine
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Patent Operations/RBW
; STREET: 1840 Behavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04681
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.

REFERENCE/DOCKET NUMBER: A-287
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 3116 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 34..2994
PCT-US95-04681-14

Query Match 100.0%; Score 12; DB 5; Length 3116;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
|||||

Db 731 CAGGAGCCTCGA 720

RESULT 10
US-08-222-616-34/C
Sequence 34, Application US/08222616
Patent No. 5635177
GENERAL INFORMATION:
APPLICANT: Bennett, Brian D.
APPLICANT: Goeddel, David
APPLICANT: Lee, James M.
APPLICANT: Matthews, William
APPLICANT: Tsai, Siao Ping
APPLICANT: Wood, William I.
TITLE OF INVENTION: PROTEIN TYROSINE KINASE AGONIST ANTIBODIES
TITLE OF INVENTION: ANTIBODIES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080

COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/222,616
FILING DATE: 4-APR-1994
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/00586
FILING DATE: 22-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/826935
FILING DATE: 22-JAN-1992
ATTORNEY/AGENT INFORMATION:
NAME: Lee, Wendy M.
REGISTRATION NUMBER:
REFERENCE/DOCKET NUMBER: 821P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1994
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 34:

SEQUENCE CHARACTERISTICS:
LENGTH: 3348 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-222-616-34

Query Match 100.0%; Score 12; DB 1; Length 3348;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
|||||

Db 698 CAGGAGCCTCGA 687

RESULT 11
PCT-US95-04228-34/C
Sequence 34, Application PC/TUS9504228
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Bennett, Brian D.
APPLICANT: Goeddel, David
APPLICANT: Lee, James M.
APPLICANT: Matthews, William
APPLICANT: Tsai, Siao Ping
APPLICANT: Wood, William I.
TITLE OF INVENTION: PROTEIN TYROSINE KINASE AGONIST ANTIBODIES
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080

COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04228
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/222616
FILING DATE: 04-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Wendy M. Lee
REGISTRATION NUMBER: 00,000
REFERENCE/DOCKET NUMBER: 821P3PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1994
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 34:

SEQUENCE CHARACTERISTICS:
LENGTH: 3348 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
PCT-US95-04228-34

Query Match 100.0%; Score 12; DB 5; Length 3348;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
|||||

Db 698 CAGGAGCCTCGA 687

RESULT 12
US-09-428-711A-13/C
Sequence 13, Application US/09428711A

```
; Patent No. 6358720
; GENERAL INFORMATION:
; APPLICANT: Muramatsu, Masaaki
; APPLICANT: Shirasawa, Takuji
; APPLICANT: Tokumitsu, Hiroshi
; APPLICANT: No. 6358720uchl, Teruhisa
; TITLE OF INVENTION: SERINE/THREONINE PROTEIN KINASE
; FILE REFERENCE: 06501-045001
; CURRENT APPLICATION NUMBER: US/09/428,711A
; CURRENT FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: PCT/JP98/01246
; PRIOR FILING DATE: 1998-03-23
; PRIOR APPLICATION NUMBER: JP 9/124798
; PRIOR FILING DATE: 1997-04-28
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 3995
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (388)...(3540)
; NAME/KEY: misc_feature
; LOCATION: (1)...(3995)
; OTHER INFORMATION: n = A,T,C or G
US-09-428-711A-13
```

```
Query Match 100.0%; Score 12; DB 4; Length 3995;
Best Local Similarity 100.0%; Pred. No. 1.4e-02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 caggagcctcga 12
|||||
Db 2875 CAGGAGCCTCGA 2864
```

```
RESULT 13
US-08-254-358-1
; Sequence 1, Application US/08254358
; Patent No. 5658785
; GENERAL INFORMATION:
; APPLICANT: Johnson, Philip R.
; TITLE OF INVENTION: Adeno-Associated Virus Materials and
; TITLE OF INVENTION: Methods
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 S. Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/254,358
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5658785and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31975
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; NAME: No. 5658785and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31975
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
```

```
; LENGTH: 4680 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-254-358-1

Query Match 100.0%; Score 12; DB 1; Length 4680;
Best Local Similarity 100.0%; Pred. No. 1.4e-02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
|||||
Db 2993 CAGGAGCCTCGA 3004

RESULT 14
US-08-475-391-1
; Sequence 1, Application US/08475391
; Patent No. 5786211
; GENERAL INFORMATION:
; APPLICANT: Johnson, Philip R.
; TITLE OF INVENTION: Adeno-Associated Virus Materials and
; TITLE OF INVENTION: Methods
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 S. Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,391
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/254,358
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5786211and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31975
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4680 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-475-391-1

Query Match 100.0%; Score 12; DB 1; Length 4680;
Best Local Similarity 100.0%; Pred. No. 1.4e-02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
|||||
Db 2993 CAGGAGCCTCGA 3004

RESULT 15
```

US-08-709-609-1
; Sequence 1, Application US/08709609
; Patent No. 5858775
; GENERAL INFORMATION:
; APPLICANT: Johnson, Phillip R.
; TITLE OF INVENTION: Adeno-Associated Virus Materials and
; TITLE OF INVENTION: Methods
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 S. Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08709,609
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5858775and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31975
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4680 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-709-609-1

Query Match 100.0%; Score 12; DB 2; Length 4680;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 caggagcctcga 12
| | | | | | | | | | | | | | | |
Db 2993 CAGGAGCCTCGA 3004

Search completed: July 31, 2002, 12:23:49
Job time: 11852 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 12:12:05 ; Search time 5855.71 Seconds
(without alignments)
27.659 Million cell updates/sec

Title: US-09-899-718A-6

Perfect score: 12

Sequence: 1 caggagcctga 12

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_esthba:*
 - 3: em_estin:*
 - 4: em_estnu:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_hic:*
 - 9: gb_estl:*
 - 10: gb_est2:*
 - 11: gb_hic:*
 - 12: gb_gss:*
 - 13: em_gss_hum:*
 - 14: em_gss_inv:*
 - 15: em_gss_pln:*
 - 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
c 1	12	100.0	109	10	BF371310	BF371310 RC6-FN011
c 2	12	100.0	110	12	A2747991	A2747991 RPCI-24-1
c 3	12	100.0	118	12	BH354246	BH354246 CH230-680
c 4	12	100.0	129	12	A2748898	A2748898 RPCI-24-1
c 5	12	100.0	133	9	AA584951	AA584951 nn41b12.s
c 6	12	100.0	144	10	B1518243	B1518243 603042090
c 7	12	100.0	145	10	BF812758	BF812758 CM2-CI018
c 8	12	100.0	153	10	BF926690	BF926690 PM1-NT023
c 9	12	100.0	172	10	BE827454	BE827454 CM2-ET001
c 10	12	100.0	173	10	BI002833	BI002833 MR3-HN016
c 11	12	100.0	176	12	A2711167	A2711167 RPCI-23-4
c 12	12	100.0	177	9	AV640236	AV640236 AV640236
c 13	12	100.0	186	9	AA038245	AA038245 mi82e03.r
c 14	12	100.0	193	9	AA074114	AA074114 zf9f05.r
c 15	12	100.0	198	9	AA657774	AA657774 nu06f04.s
c 16	12	100.0	198	9	BB592607	BB592607 BB592607
c 17	12	100.0	211	9	AA770456	AA770456 ah89f12.s

18	12	100.0	211	9	AA450935	AA450935 vf85b01.r
c 19	12	100.0	212	9	BB588724	BB588724 BB588724
c 20	12	100.0	212	9	BE091664	BE091664 IL2-BT073
c 21	12	100.0	214	10	BF605617	BF605617 272154.MA
c 22	12	100.0	215	10	BM054775	BM054775 le85c04.x
c 23	12	100.0	219	10	BG275999	BG275999 NXSL_147_
c 24	12	100.0	229	9	BB354103	BB354103 BB354103
c 25	12	100.0	229	10	BI051362	BI051362 RC5-GN028
c 26	12	100.0	229	10	D21628	D21628 MUS75D04.mo
c 27	12	100.0	230	10	BE838749	BE838749 RC6-FN011
c 28	12	100.0	231	10	BF402808	BF402808 UI-R-CA0-
c 29	12	100.0	232	10	BF399261	BF399261 UI-R-CA1-
c 30	12	100.0	233	12	BH204013	BH204013 Sm1-46G14
c 31	12	100.0	249	10	BM068805	BM068805 le85c04.y
c 32	12	100.0	252	9	AI020066	AI020066 ubl1a04.r
c 33	12	100.0	253	10	BE714112	BE714112 MR4-HT071
c 34	12	100.0	255	9	AI337414	AI337414 tb99f05.x
c 35	12	100.0	258	10	BF743579	BF743579 QVO-BT084
c 36	12	100.0	259	9	AA365282	AA365282 EST76054
c 37	12	100.0	260	9	AV735802	AV735802 AV735802
c 38	12	100.0	263	9	BE143676	BE143676 MRQ-HN016
c 39	12	100.0	265	10	BI008774	BI008774 QV3-RT006
c 40	12	100.0	270	9	AA348160	AA348160 EST54567
c 41	12	100.0	270	9	AA365578	AA365578 EST76401
c 42	12	100.0	270	9	BB590276	BB590276 BB590276
c 43	12	100.0	271	9	AI564849	AI564849 tm86a06.x
c 44	12	100.0	271	9	AW325543	AW325543 17495.MAR
c 45	12	100.0	272	10	R22443	R22443 yh23g06.sl

ALIGNMENTS

RESULT 1
BF371310/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

BF371310 109 bp mRNA linear EST 24-NOV-2000
RC6-FN0114-090800-011-E11 FN0114 Homo sapiens cDNA, mRNA sequence.
BF371310
BF371310.1 GI:11333244
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 109)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC6&t2=RC6-FN0114-
090800-011-E11&t3=2000-08-09&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 35
High quality sequence stop: 65.
Location/Qualifiers
1. 109
/organism="Homo sapiens"

```

/db_xref="taxon:9606"
/clone_lib="FN0114"
/dev_stage="Adult"
/note="Organ: prostate_normal; Vector: puc18; Site_1: SmaI
; Site_2: SmaI; A mini-library was made by cloning
products derived from ORSTES PCR (U.S. Letters Patent
application No. 196,716 - Ludwig Institute for Cancer
Research) profiles into the pUC 18 vector. Reverse
transcription of tissue mRNA and cDNA amplification were
performed under low stringency conditions."
BASE COUNT      27 a      33 c      26 g      23 t
ORIGIN

```

```

Query Match      100.0%; Score 12; DB 10; Length 109;
Best Local Similarity 100.0%; Pred. No. 8.6e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 caggagctcga 12
|||||
DB 86 CAGGAGCCTCGA 75

```

```

RESULT 2
A2747991
LOCUS      A2747991      110 bp      DNA      linear      GSS 25-JAN-2001
DEFINITION RPI-24-125D2.TV RPI-24 Mus musculus genomic clone RPI-24-125D2,
DNA sequence.
ACCESSION  A2747991
VERSION    A2747991.1 GI:12532400
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 110)
Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPI-24
Unpublished (1999)
Other_GSSs: RPI-24-125D2.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 125 row: D column: 2
Seq primer: T7
Class: BAC ends.

```

```

FEATURES
source
1..110
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPI-24-125D2"
/clone_lib="RPI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI;
RPI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."
BASE COUNT      34 a      24 c      25 g      27 t
ORIGIN

```

```

Query Match      100.0%; Score 12; DB 12; Length 110;
Best Local Similarity 100.0%; Pred. No. 8.7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 caggagctcga 12
|||||
DB 27 CAGGAGCCTCGA 38

```

```

RESULT 3
BH354246/c
LOCUS      BH354246      118 bp      DNA      linear      GSS 03-DEC-2001
DEFINITION CH230-68019.TV CHORI-230 Segment 1 Rattus norvegicus genomic clone
CH230-68019, DNA sequence.
ACCESSION  BH354246
VERSION    BH354246.1 GI:17284980
KEYWORDS   GSS.
SOURCE     Norway rat.
ORGANISM   Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 118)
Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
,A., Gebregorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
Jong,P. and Fraser,C.M.
Rat BAC End Sequences from Library CHORI-230 EcoRI segment
Unpublished (1999)
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(http://www.choi.org/bacpac/rat230.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.choi.org/bacpac/ordering_information.htm). BAC end
page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 68 row: O column: 19
Seq primer: T7
Class: BAC ends.

```

```

FEATURES
source
1..118
/organism="Rattus norvegicus"
/strain="BN/SSNHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-68019"
/clone_lib="CHORI-230 Segment 1"
/sex="Female"
/cell_type="Brain"
/note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by
Pieter de Jong"
BASE COUNT      23 a      28 c      33 g      34 t
ORIGIN

```

```

Query Match      100.0%; Score 12; DB 12; Length 118;
Best Local Similarity 100.0%; Pred. No. 8.8e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 caggagctcga 12
|||||
DB 33 CAGGAGCCTCGA 22

```

```

RESULT 4
AZ748698

```

LOCUS AZ748898 129 bp DNA linear GSS 25-JAN-2001
 DEFINITION RPCI-24-125F2-TV RPCI-24 Mus musculus genomic clone RPCI-24-125F2,
 DNA sequence.
 ACCESSION AZ748898
 VERSION AZ748898.1 GI:12534057
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 129)
 Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Akinret, B., Lewins, M.,
 Russell, D., de Jong, P. and Fraser, C.M.
 Mouse BAC End Sequences from Library RPCI-24
 Unpublished (1999)
 JOURNAL Other_GSSs: RPCI-24-125F2-TV
 COMMENT Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-24. For BAC
 library availability, please contact Pieter de Jong
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC
 Resources (http://www.choi.org/bacpac/orderingframe.html). BAC end
 page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 125 row: F column: 2
 Seq primer: T7
 Class: BAC ends.

FEATURES
 source Location/Qualifiers
 1..129
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-24-125F2"
 /clone_lib="RPCI-24"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /note="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI;
 RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
 library was cloned in the pTARBAC1 cloning vector at the
 BamHI sites using MboI partially digested male C57BL/6J
 DNA."
 BASE COUNT 38 a 29 c 29 g 33 t
 ORIGIN

Query Match 100.0%; Score 12; DB 12; Length 129;
 Best Local Similarity 100.0%; Pred. No. 9e+03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggagcctcga 12
 |||
 Db 28 CAGGAGCCTCGA 39

RESULT 5
 AA584951
 LOCUS
 DEFINITION nn41b12.s1 NCI_CGAP_GC5 Homo sapiens cDNA clone IMAGE:1086431 3',
 similar to SW:DHCA_MOUSE P48758 CARBONYL REDUCTASE ;, mRNA
 sequence.
 ACCESSION AA584951
 VERSION AA584951.1 GI:2367731
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 133)

AUTHORS
 TITLE
 JOURNAL
 COMMENT

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: Stratagene, Inc.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality
 Insert Length: 1082 Std Error: 0.00
 Seq primer: -40ml3 fwd. ET from Amersham
 High quality sequence stop: 1.
 Location/Qualifiers
 1..133
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1086431"
 /clone_lib="NCI_CGAP_GC5"
 /tissue_type="germ cell tumor"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
 cell tumors. 5' adaptor sequence: 5' GAATTCGGCACCAG 3' 3',
 adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3' Average
 insert size: 0.7 kb."
 BASE COUNT 26 a 47 c 33 g 27 t
 ORIGIN

Query Match 100.0%; Score 12; DB 9; Length 133;
 Best Local Similarity 100.0%; Pred. No. 9.1e+03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggagcctcga 12
 |||
 Db 86 CAGGAGCCTCGA 97

RESULT 6
 BI518243
 LOCUS
 DEFINITION 603042090T1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5182415 3',
 mRNA sequence.
 ACCESSION BI518243
 VERSION BI518243.1 GI:15343035
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 144)
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM11455 row: c column: 24
 High quality sequence start: 24
 High quality sequence stop: 144.

```

FEATURES
  source
    Location/Qualifiers
      1. .144
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:5182415"
        /clone_lib="NIH_MGC_116"
        /lab_host="DH10B"
        /note="Organ: pooled colon, kidney, stomach; Vector:
        pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
        source anonymous pool of 3 colons, age 26 yo male, 49 yo
        female, 71 yo male colon; 46 yo male kidney, and pool of 2
        stomachs, 62 yo male and 70 yo female. Library is
        oligo-dT primed and directionally cloned (EcoRV site is
        destroyed upon cloning). Average insert size 1.4 kb,
        insert size range 1-3 kb. Library is normalized and
        enriched for full-length clones and was constructed by C.
        Gruber (Invitrogen). Research Genetics tracking code
        023. Note: this is a NIH_MGC Library."
      31 a 36 c 33 g 44 t
      BASE COUNT
      ORIGIN

    Query Match 100.0%; Score 12; DB 10; Length 144;
    Best Local Similarity 100.0%; Pred. No. 9.3e+03;
    Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

    QY 1 caggagcctcga 12
      |||||
    Db 127 CAGGAGCCTCGA 138

  RESULT 7
  BF812758
  LOCUS CM2-CI0180-221100-540-b07 CI0180 Homo sapiens cDNA, mRNA sequence.
  DEFINITION BF812758
  ACCESSION BF812758
  VERSION BF812758.1 GI:12142691
  KEYWORDS EST.
  SOURCE human.
  ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE 1 (bases 1 to 145)
  AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
    Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
    Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
    Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
    ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
    Simpson,A.J.
  TITLE Shotgun sequencing of the human transcriptome with ORF expressed
    sequence tags
  JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE 20202663
  COMMENT Contact: Simpson A.J.G.
    Laboratory of Cancer Genetics
    Ludwig Institute for Cancer Research
    Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
    Brazil
    Tel: +55-11-2704922
    Fax: +55-11-2707001
    Email: asimpson@ludwig.org.br
    This sequence was derived from the FAPESP/LICR Human Cancer Genome
    Project. This entry can be seen in the following URL
    (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM2&t2=CM2-CI0180-
    221100-540-b07&t3=2000-11-22&t4=1)
    Seq primer: puc 18 forward
    High quality sequence start: 24
    High quality sequence stop: 144.
  FEATURES
    Location/Qualifiers
      1. .145
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone_lib="CI0180"

FEATURES
  source
    /dev_stage="Adult"
    /note="Organ: colon.ins; Vector: puc18; Site_1: SmaI;
    Site_2: SmaI; A mini-library was made by cloning products
    derived from ORESTES PCR (U.S. Letters Patent application
    No. 196,716 - Ludwig Institute for Cancer Research)
    profiles into the puc 18 vector. Reverse transcription of
    tissue mRNA and cDNA amplification were performed under
    low stringency conditions."
    31 a 35 c 36 g 43 t
    BASE COUNT
    ORIGIN

    Query Match 100.0%; Score 12; DB 10; Length 145;
    Best Local Similarity 100.0%; Pred. No. 9.3e+03;
    Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

    QY 1 caggagcctcga 12
      |||||
    Db 94 CAGGAGCCTCGA 105

  RESULT 8
  BF926690/c
  LOCUS PMI-NT0235-011200-002-a10 NT0235 Homo sapiens cDNA, mRNA sequence.
  DEFINITION BF926690
  ACCESSION BF926690
  VERSION BF926690.1 GI:12324010
  KEYWORDS EST.
  SOURCE human.
  ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE 1 (bases 1 to 153)
  AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
    Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
    Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
    Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
    ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
    Simpson,A.J.
  TITLE Shotgun sequencing of the human transcriptome with ORF expressed
    sequence tags
  JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE 20202663
  COMMENT Contact: Simpson A.J.G.
    Laboratory of Cancer Genetics
    Ludwig Institute for Cancer Research
    Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
    Brazil
    Tel: +55-11-2704922
    Fax: +55-11-2707001
    Email: asimpson@ludwig.org.br
    This sequence was derived from the FAPESP/LICR Human Cancer Genome
    Project. This entry can be seen in the following URL
    (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PMI&t2=PMI-NT0235-
    011200-002-a10&t3=2000-12-01&t4=1)
    Seq primer: puc 18 forward
    High quality sequence start: 13
    High quality sequence stop: 152.
  FEATURES
    Location/Qualifiers
      1. .153
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone_lib="NT0235"
        /dev_stage="Adult"
        /note="Organ: nervous.tumor; Vector: puc18; Site_1: SmaI;
        Site_2: SmaI; A mini-library was made by cloning products
        derived from ORESTES PCR (U.S. Letters Patent application
        No. 196,716 - Ludwig Institute for Cancer Research)
        profiles into the puc 18 vector. Reverse transcription of
        tissue mRNA and cDNA amplification were performed under
        low stringency conditions."
      32 a 40 c 50 g 31 t
      BASE COUNT
      ORIGIN

```

Query Match 100.0%; Score 12; DB 10; Length 153;
 Best Local Similarity 100.0%; Pred. No. 9.5e+03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
 |||
 Db 144 CAGGAGCCTCGA 133

RESULT 9
 BE827454
 LOCUS BE827454 172 bp mRNA linear EST 22-SEP-2000
 DEFINITION CM2-ET0016-310500-206-d05 ET0016 Homo sapiens cDNA, mRNA sequence.
 BE827454
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE EST.
 ORGANISM human.

REFERENCE
 AUTHORS
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

TITLE
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL
 MEDLINE
 COMMENT
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-CM2-ET0016-310500-206-d05&tl=2000-05-31&tl=1)

Seq primer: puc 18 forward
 High quality sequence start: 6
 High quality sequence stop: 145.
 Location/Qualifiers
 1. 172
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="ET0016"
 /dev_stage="Adult"
 /note="Organ: lung_tumor; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

FEATURES

source
 1. 172
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="ET0016"
 /dev_stage="Adult"

/note="Organ: lung_tumor; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 35 a 49 c 35 g 53 t

BASE COUNT
 ORIGIN

Query Match 100.0%; Score 12; DB 10; Length 172;
 Best Local Similarity 100.0%; Pred. No. 9.8e+03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggagcctcga 12
 |||
 Db 55 CAGGAGCCTCGA 66

RESULT 10

BI002833
 LOCUS BI002833 173 bp mRNA linear EST 13-JUN-2001
 DEFINITION MR3-HN0166-310101-002-e06 HN0166 Homo sapiens cDNA, mRNA sequence.
 BI002833
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE EST.
 ORGANISM human.

REFERENCE
 AUTHORS
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

TITLE
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL
 MEDLINE
 COMMENT
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR3&tl=2-MR3-HN0166-310101-002-e06&tl=2001-01-31&tl=1)

Seq primer: puc 18 forward
 High quality sequence start: 5
 High quality sequence stop: 173.
 Location/Qualifiers
 1. 173
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="HN0166"
 /dev_stage="Adult"
 /note="Organ: head_normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

FEATURES

source
 1. 173
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="HN0166"
 /dev_stage="Adult"

/note="Organ: head_normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 36 a 59 c 50 g 28 t

BASE COUNT
 ORIGIN

Query Match 100.0%; Score 12; DB 10; Length 173;
 Best Local Similarity 100.0%; Pred. No. 9.8e+03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
 |||
 Db 162 CAGGAGCCTCGA 173

RESULT 11

AZ271167
 LOCUS AZ271167 176 bp DNA linear GSS 26-JUL-2000
 DEFINITION RPCI-23-444J17.TJ RPCI-23 Mus musculus genomic clone RPCI-23-444J17
 AZ271167
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 DNA sequence.
 AZ271167
 GSS.
 house mouse.

```

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 176)
AUTHORS Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akimret,
B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.,
and Fraser, C.M.
TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished (1999)
COMMENT Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (inforesgen.com). BAC end page:
http://www.tigr.org/tldb/bac\_ends/mouse/bac\_end\_intro.html
Plate: 444 row: J column: 17
Seq primer: SP6
Class: BAC ends.
FEATURES
    source
        1..176
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="RPCI-23-44A117"
            /clone_lib="RPCI-23"
            /sex="Female"
            /lab_host="DH10B"
            /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
            EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
            brain genomic DNA was isolated and partially digested
            with a combination of EcoRI and EcoRI Methylase. Size
            selected DNA was cloned into the pBACe3.6 vector at the
            EcoRI sites. The ligation products were transformed into
            DH10B electrocompetent cells (BRL Life Technologies)."
        44 a 48 c 52 g 32 t
BASE COUNT 44 a 48 c 52 g 32 t
ORIGIN
    1 caggagcctcga 12
    |||||
    Db 12 CAGGAGCCTCGA 23

Query Match 100.0%; Score 12; DB 12; Length 176;
Best Local Similarity 100.0%; Pred. No. 9.8e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
|||||
Db 12 CAGGAGCCTCGA 23

RESULT 12
AV640236 177 bp mRNA linear EST 15-DEC-2000
LOCUS AV640236 Chlamydomonas reinhardtii 5% CO2
DEFINITION cDNA clone HCL012fil_r 5', mRNA sequence.
ACCESSION AV640236
VERSION AV640236.1 GI:10783564
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii.
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadales; Chlamydomonas.
REFERENCE 1 (bases 1 to 177)
AUTHORS Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K.,
Nakamura, Y., and Tabata, S.
TITLE Generation of expressed sequence tags from low-CO2 and high-CO2
adapted cells of Chlamydomonas reinhardtii
JOURNAL DNA Res. 7 (5), 305-307 (2000)
MEDLINE 20539644

Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES
    Location/Qualifiers
        1..177
            /organism="Chlamydomonas reinhardtii"
            /strain="C9"
            /db_xref="taxon:3055"
            /clone="HCL012fil_r"
            /clone_lib="Chlamydomonas reinhardtii 5% CO2"
            /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
            XhoI; The cDNA library was constructed from cells cultured
            in a medium with bubbling air containing 5% carbon
            dioxide"
        34 a 56 c 57 g 30 t
BASE COUNT 34 a 56 c 57 g 30 t
ORIGIN
    1 caggagcctcga 12
    |||||
    Db 135 CAGGAGCCTCGA 146

RESULT 13
AA038245/c 186 bp mRNA linear EST 28-AUG-1996
LOCUS AA038245 Soares mouse p3NMFl9.5 Mus musculus cDNA clone
DEFINITION IMAGE:473116 5', mRNA sequence.
ACCESSION AA038245
VERSION AA038245.1 GI:1513669
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 186)
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Willson, R. and
Waterston, R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:283860
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 180.
FEATURES
    source
        1..186
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /clone="IMAGE:473116"
            /clone_lib="Soares mouse p3NMFl9.5"
            /dev_stage="19.5 dpc total fetus"
            /lab_host="DH10B (ampicillin resistant)"
            /note="Vector: pT7T3D (Pharmacia) with a modified
            polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
            was primed with a Not I - oligo(dT) primer [5'
            TGTTACCAATCTGAGTGGGCGCGCATTTTTTTTTTTT 3'],

```

double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."

```

BASE COUNT      46 a  55 c  52 g  33 t
ORIGIN

Query Match      100.0%; Score 12; DB 9; Length 186;
Best Local Similarity 100.0%; Pred. No. 1e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
   |||
Db  69 CAGGAGCCTCGA 80

RESULT 15
AA657774      198 bp  mRNA  linear  EST 05-NOV-1997
LOCUS      nu06f04.s1 NCI_CGAP_Pr2 Homo sapiens cDNA clone IMAGE:1207231, mRNA
DEFINITION      sequence.
ACCESSION      AA657774
VERSION        AA657774.1 GI:2593928
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 198)
AUTHORS        NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE          National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
               Tumor Gene Index
JOURNAL        Unpublished (1997)
COMMENT        Contact: Robert Strausberg, Ph.D.
               Email: cgapbs-r@mail.nih.gov
               Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui, M.D.
               , Michael Emmert-Buck, M.D., Ph.D.
               cDNA Library Preparation: David B. Krizman, Ph.D.
               cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
               DNA Sequencing by: Washington University Genome Sequencing Center
               Clone distribution: NCI-CGAP clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               www-bio.llnl.gov/bbrp/image/image.html
               Seq primer: -40m13 fwd. EF from AmerSham.

FEATURES       source
               Location/Qualifiers
               1..198
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:1207231"
                /clone_lib="NCI_CGAP_Pr2"
                /sex="Male"
                /dev_stage="45 years old"
                /lab_host="DH10B"
                /note="vector: pAMP10; Site_1: NotI; Site_2: EcoRI; 1st
                strand cDNA was primed with oligo(dT)17 on 50 ng of
                DNase-treated, total cellular RNA obtained from 5,000-10
                ,000 microdissected preneoplastic cells
                histologically-determined to be prostatic intraepithelial
                neoplasia 2 (PIN2) cells. Double-stranded cDNA was
                ligated to EcoRI adaptors, 5 cycles of PCR applied to the
                cDNA with an adaptor-specific primer, and the resulting
                PCR product subcloned into pAMP10 by the UDG-cloning
                method (Life technologies). Average insert size is 600
                bp. NOTE: Not directionally cloned. This library was
                constructed by David Krizman."
BASE COUNT      67 a  60 c  47 g  24 t
ORIGIN

Query Match      100.0%; Score 12; DB 9; Length 198;
Best Local Similarity 100.0%; Pred. No. 1e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
   |||
Db  71 CAGGAGCCTCGA 82

Search completed: July 31, 2002, 12:12:09
Job time: 13482 sec

```

```

BASE COUNT      46 a  55 c  52 g  33 t
ORIGIN

Query Match      100.0%; Score 12; DB 9; Length 186;
Best Local Similarity 100.0%; Pred. No. 1e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
   |||
Db  120 CAGGAGCCTCGA 109

RESULT 14
AA074114      193 bp  mRNA  linear  EST 07-OCT-1996
LOCUS      zf79f05.r1 Soares_pineal_gland_N3HPG Homo sapiens cDNA clone
DEFINITION      IMAGE:383169 5', mRNA sequence.
ACCESSION      AA074114
VERSION        AA074114.1 GI:1614174
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 193)
AUTHORS        Hillier, J., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman
               , M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
               Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston
               , R., Williamson, A., Wohldmann, P., and Willson, R.
               The WashU-Merck EST Project
JOURNAL        Unpublished (1995)
COMMENT        Contact: Wilson RK
               Washington University School of Medicine
               4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
               Tel: 314 286 1800
               Fax: 314 286 1810
               Email: est@wustl.edu
               This clone is available royalty-free through LLNL; contact the
               IMAGE Consortium (info@image.llnl.gov) for further information.
               Seq primer: -28M13 rev2 from AmerSham
               High quality sequence stop: 177.

FEATURES       source
               Location/Qualifiers
               1..193
                /organism="Homo sapiens"
                /db_xref="GDB:1291426"
                /db_xref="taxon:9606"
                /clone="IMAGE:383169"
                /clone_lib="Soares_pineal_gland_N3HPG"
                /lab_host="DH10B (ampicillin resistant)"
                /note="Organ: pineal gland; Vector: pT73D (Pharmacia)
                with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                1st strand cDNA was primed with a Not I - oligo(dT) primer
                [5' TGTTACCATCTGAAGTGGAGCGCGCGTGTGTGTGTGTGTGTGTGT 3']
                , double-stranded cDNA was size selected, ligated to Eco
                RI adaptors (Pharmacia), digested with Not I and cloned
                into the Not I and Eco RI sites of a modified pT73 vector
                (Pharmacia). Library constructed by Bento Soares and
                M.Fatima Bonaldo."
BASE COUNT      40 a  67 c  47 g  36 t  3 others
ORIGIN

Query Match      100.0%; Score 12; DB 9; Length 193;
Best Local Similarity 100.0%; Pred. No. 1e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```


PT Transgenic animals expressing antigen reducing enzyme and complement
 PT inhibitor - used for production of materials suitable for human
 PT transplantation having a reduced risk of rejection

XX Example 1.1; Page 64; 146pp; English.

XX A sense primer (AAT63579) and antisense primer (AAT63580) were designed
 CC to amplify a 1033 bp sequence of the 3' flanking region of the
 CC human CD59 gene, with the introduction of a 5' SacI site and 3'
 CC XhoI site. The PCR product was joined to CD59 cDNA contg. exons 2,
 CC 3 and 4 plus 667 bp of 3' untranslated region (see also AAT63577-78)
 CC to produce a CD59 minigene. This minigene was shown to retain
 CC biological function when expressed in transgenic pigs and mice.
 CC Expression of a complement inhibitor such as CD59 on endothelial
 CC cells of transgenic animals can provide materials suitable for
 CC transplantation to humans, suppressing complement activation and
 CC reducing immune reaction.

XX Sequence 32 BP; 3 A; 9 C; 11 G; 9 T; 0 other;

Query Match 100.0%; Score 12; DB 18; Length 32;

Best Local Similarity 100.0%; Pred. No. 8.9e+02;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12

|||||

Db 17 CAGGAGCCTCGA 6

RESULT 2

ABA03059/c

ID ABA03059 standard; DNA; 58 BP.

XX ABA03059;

DT 05-FEB-2002 (first entry)

DE Oligonucleotide #2 used to construct a HA fusion sequence.

XX Human; serum albumin; HA; antiinflammatory; immunosuppressive; cardiac;
 KW neoptotic; neuroprotective; gene therapy; immune disorder; wound healing;
 KW hyperproliferative disorder; renal disorder; cardiovascular disorder;
 KW respiratory disorder; neurological disease; endocrine disorder;
 KW reproductive system disorder; infectious disease;
 KW gastrointestinal disorder; ss.

XX Homo sapiens.

XX OS

XX PN WO200179444-A2.

XX XX

XX PD 25-OCT-2001.

XX XX

XX PF 12-APR-2001; 2001WO-US12013.

XX XX

XX PR 12-APR-2000; 2000US-229358P.

XX PR 25-APR-2000; 2000US-199384P.

XX PR 21-DEC-2000; 2000US-256931P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX PA

XX PI Rosen CA, Haseltine WA;

XX XX

XX DR WPI; 2001-616755/71.

XX XX

XX Albumin fusion proteins comprising a therapeutic protein and albumin,
 PT useful in the treating immune system disorders (e.g. transplant
 PT rejection), blood related disorders (e.g. myocardial infarction) and
 PT hyperproliferative disorders -

XX Example 2; Page 465; 606pp; English.

XX The present invention relates to albumin fusion proteins, which comprise

CC a therapeutic protein and albumin. The albumin fusion proteins are useful
 CC in the treatment, prevention, diagnosis, and/or detection of
 CC diseases/disorders such as immune system disorders (e.g. transplant
 CC rejection), blood related disorders (e.g. myocardial infarction),
 CC hyperproliferative disorders (e.g. childhood acute myeloid leukemia),
 CC renal disorders (e.g. glomerulonephritis), cardiovascular disorders (e.g.
 CC arrhythmias), respiratory disorders (e.g. non-allergic rhinitis),
 CC neurological diseases (e.g. Alzheimer's disease), endocrine disorders
 CC (e.g. pheochromocytoma), reproductive system disorders (e.g. syphilis),
 CC infectious diseases (e.g. measles), gastrointestinal disorders (e.g.
 CC irritable bowel syndrome) and wound healing. In the present invention,
 CC human serum albumin (HA; see AAM52567) was used to generate fusion
 CC proteins. The present sequence was used to illustrate the invention.

XX Sequence 58 BP; 13 A; 15 C; 12 G; 18 T; 0 other;

Query Match 100.0%; Score 12; DB 22; Length 58;

Best Local Similarity 100.0%; Pred. No. 8.9e+02;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12

|||||

Db 51 CAGGAGCCTCGA 40

RESULT 3

AAD22289/c

ID AAD22289 standard; DNA; 59 BP.

XX AAD22289;

XX AC

XX DT 12-FEB-2002 (first entry)

XX XX

DE Human albumin (HA) mutagenic PCR primer #2.

XX Human; albumin; HA; fusion protein; immune system disorder; syphilis;
 KW transplant rejection; blood related disorder; myocardial infarction;
 KW hyperproliferative disorder; acute myeloid leukaemia; renal disorder;
 KW glomerulonephritis; cardiovascular disease; arrhythmia; rhinitis;
 KW respiratory disorder; neurological disease; Alzheimer's disease;
 KW endocrine disorder; pheochromocytoma; reproductive system disorder;
 KW measles; gastrointestinal disorder; irritable bowel syndrome; HIV;
 KW human immunodeficiency virus; wound healing; renal cell carcinoma;
 KW melanoma; gene therapy; PCR primer; ss.

XX Homo sapiens.

XX OS

XX PN WO200179258-A1.

XX XX

XX PD 25-OCT-2001.

XX XX

XX PF 12-APR-2001; 2001WO-US12008.

XX XX

XX PR 12-APR-2000; 2000US-229358P.

XX PR 25-APR-2000; 2000US-199384P.

XX PR 21-DEC-2000; 2000US-256931P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX PA (PRIN-) PRINCIPAL PHARM CORP.

XX XX

XX PI Rosen CA, Sadeghi H, Prior CP, Turner AJ;

XX XX

XX DR WPI; 2001-602931/68.

XX XX

XX Albumin fusion proteins comprising a therapeutic protein and albumin,
 PT useful in the treating metastatic renal cell carcinoma, metastatic
 PT melanoma, malignant melanoma, renal cell carcinoma, HIV (human
 PT immunodeficiency virus) or infection -

XX Example 2; Page 190; 325pp; English.

XX The invention relates to albumin fusion proteins comprising therapeutic

CC protein and human albumin (HA). The albumin fusion proteins are useful
 CC in the treatment, prevention, diagnosis, and/or detection of diseases,
 CC disorders such as immune system disorders (transplant rejection); blood
 CC related disorders (myocardial infarction); hyperproliferative disorders
 CC (childhood acute myeloid leukaemia); renal disorder (glomerulonephritis);
 CC cardiovascular disorders (arrhythmias); respiratory disorders
 CC (non-allergic rhinitis); neurological diseases (Alzheimer's disease);
 CC endocrine disorders (pheochromocytoma); reproductive system disorders
 CC (syphilis); infectious diseases (measles); gastrointestinal disorders
 CC (irritable bowel syndrome) and wound healing. The albumin fusion
 CC proteins are also used in the treatment of metastatic renal cell
 CC carcinoma, metastatic melanoma, malignant melanoma and HIV (human
 CC immunodeficiency virus) infection. Nucleic acid encoding albumin fusion
 CC protein is useful in gene therapy. The present sequence is a mutagenic
 CC PCR primer used for engineering the Xho and Cla I restriction sites
 CC into the DNA encoding HA protein in pPC0006 plasmid.
 CC Note: The present sequence shown in sequence listing of the specification
 CC lacks a nucleotide base at its 3' end.
 CC
 SQ Sequence 59 BP; 13 A; 16 C; 12 G; 18 T; 0 other;

Query Match 100.0%; Score 12; DB 22; Length 59;
 Best Local Similarity 100.0%; Pred. No. 8.9e+02;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggagcctcga 12
 |||||
 Db 51 CAGGAGCCTCGA 40

RESULT 4
 AAD21640/c
 ID AAD21640 standard; DNA; 59 BP.
 XX
 AC AAD21640;

DT 28-JAN-2002 (first entry)

XX Human albumin (HA) PCR primer #2.

XX Human; albumin; HA; fusion protein; therapeutic protein; vulnery;
 KW immune system disorder; transplant rejection; blood related disorder;
 KW myocardial infarction; hyperproliferative disorder; glomerulonephritis;
 KW childhood acute myeloid leukemia; cardiovascular disorder; arrhythmia;
 KW respiratory disorder; gene therapy; non-allergic rhinitis; neutropic;
 KW neurological disease; Alzheimer's disease; reproductive system disorder;
 KW endocrine disorder; pheochromocytoma; infectious disease; antiarrhythmic;
 KW measles; gastrointestinal disorder; irritable bowel syndrome; syphilis;
 KW wound healing; antiinflammatory; immunosuppressive; neuroprotective;
 KW cardiant; cytostatic; antileukaemic; antirheumatic; antimicrobial;
 KW renal disorder; PCR primer; ss.

OS Homo sapiens.

XX W0200179443-A2.

XX 25-OCT-2001.

XX 12-APR-2001; 2001WO-US11924.

XX 12-APR-2000; 2000US-229358P.

XX 25-APR-2000; 2000US-199384P.

XX 21-DEC-2000; 2000US-256931P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Haseltine WA;

XX WPI; 2001-616754/71.

XX Albumin fusion proteins comprising a therapeutic protein and albumin,
 PT useful in the treating immune system disorders (e.g. transplant

PT rejection), blood related disorders (e.g. myocardial infarction) and
 PT hyperproliferative disorders -
 XX
 PS Example 2; Page 231; 380pp; English.

XX The invention relates to albumin fusion proteins comprising therapeutic
 CC protein and human albumin (HA). Therapeutic protein fused to albumin
 CC have an extended shelf-life. The albumin fusion proteins are useful in
 CC the treatment, prevention, diagnosis and/or detection of diseases,
 CC disorders such as immune system disorders (e.g. transplant rejection),
 CC blood related disorders (e.g. myocardial infarction), hyperproliferative
 CC disorders (e.g. childhood acute myeloid leukaemia), renal disorders
 CC (e.g. glomerulonephritis), cardiovascular disorders (e.g. arrhythmias),
 CC respiratory disorders (e.g. non-allergic rhinitis), neurological
 CC diseases (e.g. Alzheimer's disease), endocrine disorders (e.g.
 CC pheochromocytoma), reproductive system disorders (e.g. syphilis),
 CC infectious diseases (e.g. measles), gastrointestinal disorders (e.g.
 CC irritable bowel syndrome) and wound healing. Nucleic acids encoding
 CC albumin fusion protein is used in gene therapy. The present sequence
 CC is a PCR primer used for engineering the Xho and Cla I restriction
 CC sites into the DNA encoding HA protein in pPC0006 plasmid.
 CC Note: This sequence SEQ.ID.NO.20 is stated to be similar to the
 CC sequence shown in the sequence listing. However this sequence
 CC contains an additional base 'C' at the 3' end which is absent in
 CC the sequence shown in sequence listing.
 XX

SQ Sequence 59 BP; 13 A; 16 C; 12 G; 18 T; 0 other;

Query Match 100.0%; Score 12; DB 22; Length 59;
 Best Local Similarity 100.0%; Pred. No. 8.9e+02;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggagcctcga 12
 |||||
 Db 51 CAGGAGCCTCGA 40

RESULT 5
 AAD20007/c
 ID AAD20007 standard; DNA; 59 BP.

XX AAD20007;

DT 18-DEC-2001 (first entry)

XX Human albumin (HA) PCR primer #2.

XX Human; albumin; HA; immune system disorder; transplant rejection;
 KW blood related disorder; myocardial infarction; glomerulonephritis;
 KW hyperproliferative disorder; childhood acute myeloid leukaemia;
 KW renal cell carcinoma; cardiovascular disorder; antimicrobial; vulnery;
 KW arrhythmia; respiratory disorder; non-allergic rhinitis; antileukaemic;
 KW neurological disease; Alzheimer's disease; endocrine disorder; measles;
 KW pheochromocytoma; reproductive system disorder; neuroprotective; syphilis;
 KW infectious disease; gastrointestinal disorder; wound healing; neutropic;
 KW irritable bowel syndrome; HIV; human immunodeficiency virus infection;
 KW cytostatic; antiinflammatory; gene therapy; immunosuppressive; cardiant;
 KW antiarthritic; antirheumatic; renal disorder; melanoma; PCR primer; ss.

OS Homo sapiens.

XX W0200179480-A1.

XX 25-OCT-2001.

XX 12-APR-2001; 2001WO-US11991.

XX 12-APR-2000; 2000US-229358P.

XX 25-APR-2000; 2000US-199384P.

XX 21-DEC-2000; 2000US-256931P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Haseltine WA;
 PI WPI; 2001-616756/71.
 XX
 XX Albumin fusion proteins comprising a therapeutic protein and albumin,
 PT useful in the treating metastatic renal cell carcinoma, metastatic
 PT melanoma, malignant melanoma, renal cell carcinoma, HIV (human
 PT immunodeficiency virus) or infection -
 XX
 XX Example 2; Page 244; 394pp; English.
 XX
 XX The invention relates to human albumin (HA) fusion proteins and their
 CC corresponding nucleic acid sequences. Therapeutic proteins fused to
 CC albumin or its fragments have an extended shelf-life. The albumin
 CC fusion proteins are useful in the treatment, prevention, diagnosis,
 CC and/or detection of diseases, disorders such as immune system
 CC disorders (e.g. transplant rejection), blood related disorders (e.g.
 CC myocardial infarction), hyperproliferative disorders (e.g. childhood
 CC acute myeloid leukaemia, metastatic renal cell carcinoma, metastatic
 CC melanoma, malignant melanoma, renal cell carcinoma), renal disorders
 CC (e.g. glomerulonephritis), cardiovascular disorders (e.g. arrhythmias),
 CC respiratory disorders (e.g. non-allergic rhinitis), neurological
 CC diseases (e.g. Alzheimer's disease), endocrine disorders (e.g.
 CC pheochromocytoma), reproductive system disorders (e.g. syphilis),
 CC infectious diseases (e.g. measles), gastrointestinal disorders (e.g.
 CC irritable bowel syndrome), HIV (human immunodeficiency virus) infection
 CC and wound healing. Nucleic acids encoding albumin fusion protein is
 CC used in gene therapy. The present sequence is a mutagenic PCR primer
 CC used for engineering the Xho and Cla I restriction sites into the
 CC fusion leader sequence just 5' end of the DNA encoding the HA
 CC protein in pPPC0006 plasmid.
 CC Note: This sequence SEQ.ID.NO.20 is stated to be similar to the
 CC sequence shown in the sequence listing. However this sequence
 CC contains an additional base 'C' at the 3' end which is absent in
 CC the sequence shown in sequence listing..
 XX
 XX Sequence 59 BP; 13 A; 16 C; 12 G; 18 T; 0 other;
 SQ

Query Match 100.0%; Score 12; DB 22; Length 59;
 Best Local Similarity 100.0%; Pred. No. 8.9e+02; Mismatches 0; Gaps 0;
 Matches 12; Conservative 0; Indels 0; Gaps 0;

Qy 1 caggagcctoga 12
 |||||
 Db 51 CAGGAGCCTCGA 40

RESULT 6
 AAK51355
 ID AAK51355 standard; DNA; 193 BP.
 XX
 XX AAK51355;
 AC
 XX 06-NOV-2001 (first entry)
 DT
 XX Human bone marrow expressed single exon probe SEQ ID NO: 25912.
 DE
 XX Human: bone marrow expressed exon; gene expression analysis; probe;
 KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
 KW
 XX Homo sapiens.
 OS
 XX WO200157276-A2.
 PN
 XX 09-AUG-2001.
 PD
 XX 30-JAN-2001; 2001WO-US00668.
 PF
 XX 04-FEB-2000; 2000US-0180312.
 PR
 XX 25-MAY-2000; 2000US-0207456.
 PR
 XX 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI; 2001-488900/53.
 XX
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human bone marrow -
 PT
 XX Example 4; SEQ ID NO: 25912; 658pp + Sequence Listing; English.
 PS
 XX The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
 CC the probes of the invention.
 XX
 XX Sequence 193 BP; 32 A; 47 C; 27 G; 87 T; 0 other;
 SQ

Query Match 100.0%; Score 12; DB 22; Length 193;
 Best Local Similarity 100.0%; Pred. No. 8.8e+02; Mismatches 0; Gaps 0;
 Matches 12; Conservative 0; Indels 0; Gaps 0;

Qy 1 caggagcctoga 12
 |||||
 Db 63 caggagcctoga 74

RESULT 7
 ABL27525
 ID ABL27525 standard; DNA; 261 BP.
 XX
 XX ABL27525;
 AC
 XX 26-MAR-2002 (first entry)
 DT
 XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 34048.
 DE
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 KW
 XX Drosophila melanogaster.
 OS
 XX WO200171042-A2.
 PN
 XX 27-SEP-2001.
 PD
 XX 23-MAR-2001; 2001WO-US09231.
 PF
 XX 23-MAR-2000; 2000US-191637P.
 PR
 XX 11-JUL-2000; 2000US-0614150.
 PR
 XX (PEKE) PE CORP NY.
 PA
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI; 2001-656860/75.
 DR
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PT
 XX Claim 1; SEQ ID NO 34048; 21pp + Sequence Listing; English.
 PS
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB157737-AB172072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 261 BP; 83 A; 66 C; 68 G; 44 T; 0 other;

Query Match 100.0%; Score 12; DB 23; Length 261;
Best Local Similarity 100.0%; Pred. No. 8.8e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
|||||

Db 109 caggagcctcga 120
|||||

RESULT 8
AAC31784/C
ID AAC31784 standard; cDNA; 292 BP.

AC AAC31784;

XX 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 35859.

DE Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.

OS Homo sapiens.

XX EP1033401-A2.

PN 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
PS Claim 1; SEQ ID 35859; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.

XX Sequence 292 BP; 63 A; 76 C; 73 G; 80 T; 0 other;

Query Match 100.0%; Score 12; DB 21; Length 292;
Best Local Similarity 100.0%; Pred. No. 8.8e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
|||||

Db 222 CAGGAGCCTCGA 211
|||||

RESULT 9
AAA69618/C

ID AAA69618 standard; cDNA; 315 BP.

XX AAA69618;

XX 08-NOV-2000 (first entry)

XX Eucalyptus grandis NADPH cDNA SEQ ID NO:92.

XX Eucalyptus grandis; Pinus radiata; modification; isoprenoid; plant;
KW metabolism; isoprenoid biosynthetic pathway; terpenoid; steroid;
KW genome mapping; physical mapping; positional cloning; forestry;
KW agriculture; medicine; fermentation; plant development; pest resistance;
KW pinene; myrcene; Monterey pine; ss.

OS Eucalyptus grandis.

XX WO200036081-A2.

XX 22-JUN-2000.

XX 16-DEC-1999; 99WO-NZ00219.

XX 17-DEC-1998; 98US-0215504.

XX 29-JUL-1999; 99US-0146441.

XX (GENE-) GENESIS RES & DEV CORP LTD.

XX (FLET-) FLETCHER CHALLENGE FORESTS LTD.

XX Havukkala IJ;

XX WPI; 2000-431575/37.

XX P-PSDB; AAB18018.

XX New plant polynucleotides encoding polypeptides involved in the
PT production and modification of isoprenoids, useful in forestry and
PT agriculture for manipulation of isoprenoid metabolism -

XX Claim 1; Page 82; 164pp; English.

XX The present invention describes plant polynucleotides encoding
CC polypeptides involved in the production and modification of isoprenoids,
CC such as terpenoid and steroid compounds. The polynucleotides are used
CC in genome mapping, in physical mapping and in positional cloning of
CC genes. The polynucleotides and polypeptides are useful in forestry and
CC agriculture for manipulation of isoprenoid metabolism, in medicine for
CC therapeutic effects, including direct application in diseased organisms
CC or indirect application by transgenic organisms and in fermentation and
CC chemical processing industries involving isoprenoids. In plant
CC applications, manipulating isoprenoid pathways or isoprenoid composition
CC may, for example, affect plant development, pest resistance, and the
CC value of extractives (e.g. pinene and myrcene). The ubiquitous and
CC varied roles of isoprenoids make the polynucleotides attractive targets
CC for biotechnical applications in a variety of fields. AAA69527 to
CC AAA69690 and AAB18004 to AAB18143 represent Eucalyptus grandis and Pinus
CC radiata polynucleotides and proteins used in the exemplification of the
CC present invention.

XX Sequence 315 BP; 77 A; 96 C; 90 G; 52 T; 0 other;

Query Match 100.0%; Score 12; DB 21; Length 315;
 Best Local Similarity 100.0%; Pred. No. 8.8e+02;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
 |||||
 DB 120 CAGGAGCCTCGA 109

RESULT 10
 AAC75741/c
 ID AAC75741 standard; cDNA; 341 BP.
 XX
 AC AAC75741;
 XX
 DT
 XX
 DE 08-FEB-2001 (first entry)
 XX Human ORFX ORF1296 polynucleotide sequence SEQ ID NO:2591.
 XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 XX vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.
 XX Homo sapiens.
 OS
 XX
 PN WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PE 31-MAR-2000; 2000WO-0508621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 XX (CURA-) CURAGEN CORP.
 PA
 XX Shinkets RA, Leach M;
 PI WPI; 2000-602362/57.
 XX P-PSDB; AAB41532.
 XX
 XX Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 XX Claim 5; Page 1852; 5507pp; English.
 PS
 XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
 CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antithyroid; and antianaemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,

CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus, lupus
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX
 SQ Sequence 341 BP; 74 A; 82 C; 125 G; 60 T; 0 other;

Query Match 100.0%; Score 12; DB 21; Length 341;
 Best Local Similarity 100.0%; Pred. No. 8.8e+02;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
 |||||
 DB 267 CAGGAGCCTCGA 256

RESULT 11
 AAC04396
 ID AAC04396 standard; cDNA; 394 BP.
 XX
 AC AAC04396;
 XX
 DT 06-OCT-2000 (first entry)
 XX Human secreted protein 5' EST, SEQ ID NO: 8471.
 DE
 XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.
 KW
 XX Homo sapiens.
 OS
 XX
 PN EP1033401-A2.
 XX
 PD 06-SEP-2000.
 XX
 PE 21-FEB-2000; 2000EP-0200610.
 XX
 PR 26-FEB-1999; 99US-0122487.
 XX
 XX (GEST) GENSET.
 FA
 XX Dumas Milne Edwards J, Duclert A, Giordano J;
 PI WPI; 2000-500381/45.
 XX
 XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 XX Claim 1; SEQ ID 8471; 71pp + CD-ROM; English.
 PS
 XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.
 XX
 SQ Sequence 394 BP; 86 A; 110 C; 122 G; 73 T; 3 other;

Query Match 100.0%; Score 12; DB 21; Length 394;
 Best Local Similarity 100.0%; Pred. No. 8.8e+02;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
 |||||

Db 117 caggagcctcga 128

RESULT 12
 AAC0874/c
 ID AAC0874 standard; cDNA; 400 BP.
 AC AAC0874;
 XX
 XX 06-OCT-2000 (first entry)
 DE Human secreted protein 5' EST, SEQ ID NO: 12949.
 XX
 XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.
 KW Homo sapiens.
 OS
 XX EP1033401-A2.
 PN
 XX
 XX 06-SEP-2000.
 PD
 XX
 XX 21-FEB-2000; 2000EP-0200610.
 PF
 XX
 XX 26-FEB-1999; 99US-0122487.
 PR
 XX (GENSET) GENSET.
 PA
 XX Dumas Milne Edwards J, Duclert A, Giordano J;
 PI WPI; 2000-500381/45.
 DR
 XX
 XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 XX Claim 1; SEQ ID 12949; 71pp + CD-ROM; English.
 PS
 XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.
 XX
 XX Sequence 400 BP; 113 A; 137 C; 106 G; 43 T; 1 other;

Query Match 100.0%; Score 12; DB 21; Length 400;
 Best Local Similarity 100.0%; Pred. No. 8.8e+02;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
 |||||

Db 49 CAGGAGCCTCGA 38

RESULT 13

AAT74033
 ID AAT74033 standard; cDNA; 450 BP.
 XX
 AC AAT74033;
 XX
 XX 15-SEP-1997 (first entry)
 DT
 XX
 DE Aloe arborescens phosphoenolpyruvate carboxylase probe.
 XX
 XX PEPCase; monocotyledonous CAM plant; carbon dioxide fixation;
 KW transformation; polymerase chain reaction; ss.
 KW
 XX Aloe arborescens.
 OS
 XX JP09107975-A.
 PN
 XX 28-APR-1997.
 PD
 XX
 XX 19-OCT-1995; 95JP-0294986.
 PF
 XX
 XX 19-OCT-1995; 95JP-0294986.
 PR
 XX (MITK) MITSUI TOATSU CHEM INC.
 PA
 XX WPI; 1997-292468/27.
 DR
 XX
 XX Phospho:enol:pyruvate carboxylase gene of a monocotyledonous CAM
 PT plant - used to transform plants so that they can fix carbon
 PT di:oxide in dry conditions
 PT
 XX Example 1; Page 10; 12pp; Japanese.
 PS
 XX The present sequence represents a probe for phosphoenolpyruvate
 CC carboxylase (PEPCase) from Aloe arborescens. A plasmid containing
 CC the DNA sequence of PEPCase can be introduced into a monocotyledonous
 CC plant to enhance PEPCase activity. The transformed plant has a high
 CC ability to fix carbon dioxide under dry conditions. It can also
 CC increase corn productivity when introduced into a corn plant.
 XX
 XX Sequence 450 BP; 96 A; 101 C; 147 G; 106 T; 0 other;

Query Match 100.0%; Score 12; DB 18; Length 450;
 Best Local Similarity 100.0%; Pred. No. 8.8e+02;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
 |||||

Db 424 caggagcctcga 435

RESULT 14
 AAZ40844/c
 ID AAZ40844 standard; DNA; 475 BP.
 XX
 AC AAZ40844;
 XX
 DT 18-JAN-2000 (first entry)
 DT
 XX
 DE Secreted protein EST coding sequence 78-6-2-B10-FL1.
 XX
 XX Secreted protein; fingerprint identification technique;
 KW chromosome mapping; human; hereditary disease; diagnosis; cancer;
 KW hyperlipidemia; cardiovascular; neurodegenerative disorder; therapy;
 KW autoimmune disease; rheumatic disease; embryogenic disorder; myopathy;
 KW renal injury; amino aciduria; hypoglycaemia; male rat infertility;
 KW hypertension; ss.
 KW
 XX Homo sapiens.
 OS
 XX WO9940189-A2.
 PN
 XX 12-AUG-1999.
 PD

```

XX 09-FEB-1999; 99WO-IB00282.
XX 09-FEB-1998; 98US-0074121.
XX 13-APR-1998; 98US-0081563.
XX 10-AUG-1998; 98US-0096116.
XX 04-SEP-1998; 98US-0099273.
XX (GEST ) GENSET.
XX Bouqueleret L, Duclert A, Dumas Milne Edwards J;
XX WPI; 1999-600966/51.
XX P-PSDB; AAY59716.
XX Extended cDNAs useful for expressing secreted proteins and to obtain
XX specific antibodies -
XX Claim 1; Page 224-225; 244pp; English.
XX This sequence encodes a human secreted protein of the invention. The
XX extended cDNAs (or genomic DNAs obtainable from them) may be used to
XX prepare PCR primers and probes. These are useful for forensic matching or
XX positive identification by DNA sequencing. They may also be used in
XX alternative fingerprint identification techniques. Antibodies against the
XX proteins encoded by the extended cDNAs are useful in identification of
XX tissue types or cell species, as well as identifying tissue specific
XX soluble proteins. The sequences can be used for chromosome mapping and
XX identification of genes associated with hereditary diseases or drug
XX response. Signal sequences from the cDNAs can be used in construction of
XX secretion vectors. Other sequences derived from the extended cDNAs can be
XX used to clone upstream genomic DNA sequences including promoters. This is
XX in turn useful for identifying proteins that interact with promoter
XX sequences. Some of the proteins may be useful in diagnosing and treating
XX several disorders including, but not limited to: cancer, hyperlipidaemia,
XX cardiovascular and neurodegenerative disorders, autoimmune diseases, and
XX rheumatic diseases, embryogenic disorders, hypertension, renal injury,
XX amino acidurias, hypoglycaemia, male rat infertility and myopathies.
XX Sequence 475 BP; 121 A; 118 C; 112 G; 121 T; 3 other;

Query Match 100.0%; Score 12; DB 20; Length 475;
Best Local Similarity 100.0%; Pred. No. 8.8e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
Db |||||
223 CAGGAGCCTCGA 212

RESULT 15
AAH97774/C
ID AAH97774 standard; DNA; 552 BP.
XX AC
XX AAH97774;
XX 10-OCT-2001 (first entry)
XX Murine 7-transmembrane G-protein coupled receptor coding sequence #18.
XX Murine; hematopoietic stem cell; signalling; vaccine; 7TM-GPCR;
XX 7-transmembrane G-protein coupled protein receptor; ds.
XX Mus sp.
XX WO200160999-A1.
XX 23-AUG-2001.
XX 14-FEB-2001; 2001WO-US04700.
XX 14-FEB-2000; 2000US-0182377.

```

```

XX (JMCL-) IMCLONE SYSTEMS INC.
XX (UYPR-) UNIV PRINCETON.
XX Lemischka IR, Witte L, Pereira DS;
XX WPI; 2001-522596/57.
XX DNA Sequences encoding 7-transmembrane G-protein coupled protein
XX receptors characteristic of hematopoietic stem cells, useful for
XX treating leukemia -
XX Claim 1; Page 30-31; 176pp; English.
XX The present invention relates to murine coding sequences for
XX 7-transmembrane G-protein coupled protein receptors (7TM-GPCRs). The
XX present sequence is one such murine 7TM-GPCR coding sequence. The present
XX sequence was derived from hematopoietic stem cells. The present sequence
XX and its corresponding protein are useful in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate 7TM-GPCR expression.
XX 7TM-GPCRs identify specific signalling molecules, to activate an
XX effector-signalling cascade that triggers an intracellular response and
XX eventually a biological effect.
XX Sequence 552 BP; 126 A; 120 C; 125 G; 95 T; 86 other;

Query Match 100.0%; Score 12; DB 22; Length 552;
Best Local Similarity 100.0%; Pred. No. 8.8e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
Db |||||
130 CAGGAGCCTCGA 119

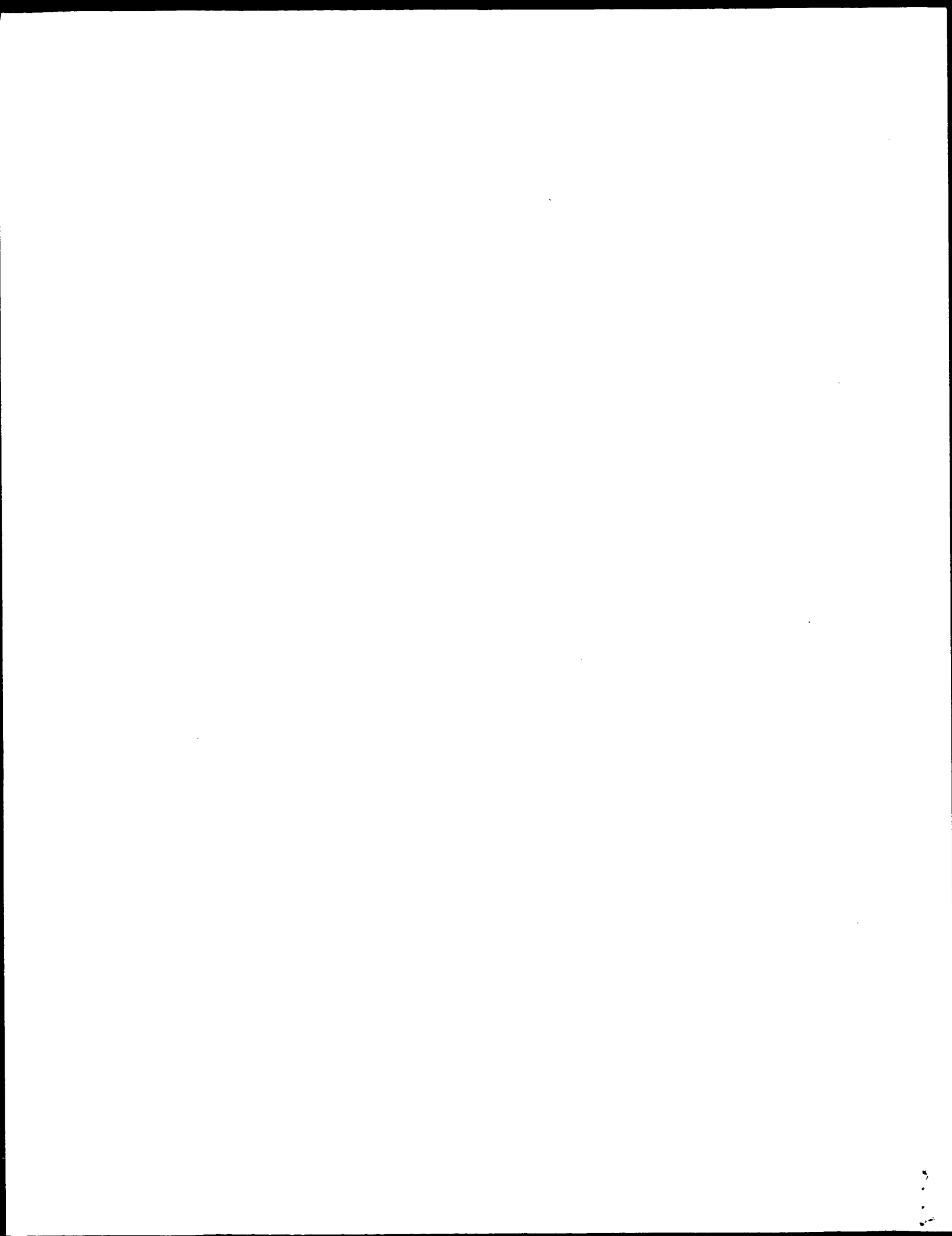
Search completed: July 31, 2002, 14:11:24
Job time: 17381 sec

```


Thu Aug 1 08:30:04 2002

us-09-899-718a-6.rng

Page 9



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 14:01:53 ; Search time 6034.22 Seconds
(without alignments)
41.616 Million cell updates/sec

Title: US-09-899-718A-6
Perfect score: 12
Sequence: 1 caggagcctcga 12

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*
2: gb_hg.*
3: gb_in.*
4: gb_cm.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pi.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

RESULT 1

AX349068

LOCUS

DEFINITION

AX349068

ACCESSION

VERSION

AX349068.1

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1. 12

/organism="synthetic construct"

/db_xref="taxon:32630"

/note="Oligonucleotide"

3 a 4 c 4 g 1 t

BASE COUNT

ORIGIN

ALIGNMENTS

AX349068 Sequence 6 from Patent WO0202785. 12 bp DNA linear PAT 06-FEB-2002

AX349068

AX349068.1

GI:18615103

synthetic construct.

synthetic construct

artificial sequence.

1 (sites)

Sprunck, S., Kluth, A., Becker, D., Luetticke, S. and Loerz, H.

Promoters of gene expression in plant caryopses

Patent: WO 0202785-A 6 10-JAN-2002;

Aventis CropScience GmbH (DE)

Location/Qualifiers

1. 12

/organism="synthetic construct"

/db_xref="taxon:32630"

/note="Oligonucleotide"

3 a 4 c 4 g 1 t

1	12	100.0	12	6	AX349068	AX349068 Sequence
2	12	100.0	218	5	AF150645	AF150645 Petrolyx
3	12	100.0	218	5	AF150652	AF150652 Mastacemb
c 4	12	100.0	258	8	TSITSAA1	AF084320 Tamarix s
5	12	100.0	442	11	G49233	G49233 D1ErtD754e
6	12	100.0	450	6	E12962	E12962 cDNA encodi
7	12	100.0	453	9	HS301F5	279832 H.sapiens c
8	12	100.0	453	9	HS302F12	279864 H.sapiens c
9	12	100.0	475	6	AR177408	AR177408 Sequence
c 10	12	100.0	498	6	AB02081S28	AB020839 Homo sapi
11	12	100.0	598	11	G39052	G39052 z11764 Zebr
12	12	100.0	680	1	AF321195	AF321195 Unculture
c 13	12	100.0	727	9	HUMSVB1A1	M36196 Human synap
14	12	100.0	761	11	G41958	G41958 SHGC-68982
c 15	12	100.0	822	6	A70150	A70150 Sequence 11
c 16	12	100.0	822	8	MAZ93109	Z93109 M.acuminata
c 17	12	100.0	903	33	AC083094	AC083094 Giardia i
c 18	12	100.0	908	8	HBHMR2M	X54658 H.brasili
c 19	12	100.0	936	33	AC076609	AC076609 Giardia i
c 20	12	100.0	939	33	AC062911	AC062911 Giardia i
c 21	12	100.0	941	33	AC085805	AC085805 Giardia i
c 22	12	100.0	942	33	AC038729	AC038729 Giardia i
c 23	12	100.0	954	33	AC067176	AC067176 Giardia i
c 24	12	100.0	959	33	AC045429	AC045429 Giardia i
c 25	12	100.0	962	33	AC071714	AC071714 Giardia i
c 26	12	100.0	966	33	AC066979	AC066979 Giardia i
c 27	12	100.0	985	33	AC085377	AC085377 Giardia i
c 28	12	100.0	987	33	AC043717	AC043717 Giardia i
c 29	12	100.0	990	4	OVU53517	U53517 Odocoileus
30	12	100.0	991	4	OHU53519	U53519 Odocoileus
31	12	100.0	1000	33	AC043716	AC043716 Giardia i
c 32	12	100.0	1000	33	AC086493	AC086493 Giardia i
c 33	12	100.0	1027	33	AC060381	AC060381 Giardia i
c 34	12	100.0	1027	33	AC065036	AC065036 Giardia i
c 35	12	100.0	1034	33	AC054763	AC054763 Giardia i
36	12	100.0	1042	33	AC047036	AC047036 Giardia i
37	12	100.0	1050	33	AC054357	AC054357 Giardia i
38	12	100.0	1056	33	AC059111	AC059111 Giardia i
39	12	100.0	1060	33	AC057542	AC057542 Giardia i
40	12	100.0	1105	33	AC055306	AC055306 Giardia i
41	12	100.0	1236	4	AY057834	AY057834 Crocidura
c 42	12	100.0	1440	4	OCU21015	U101015 Oryctolagus
43	12	100.0	1542	9	BC012857	BC012857 Homo sapi
44	12	100.0	1580	9	BC009052	BC009052 Homo sapi
c 45	12	100.0	1816	10	BC021488	BC021488 Mus muscu

```

Query Match          100.0%; Score 12; DB 6; Length 12;
Best Local Similarity 100.0%; Pred. No. 7.9e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
|||||
Db 1 CAGGAGCCTCGA 12

RESULT 2
AF150645          218 bp DNA linear VRT 22-AUG-2000
LOCUS             Petrotux sanguineus 28S ribosomal RNA gene, partial sequence.
ACCESSION         AF150645
VERSION           AF150645.1 GI:4960098
KEYWORDS          .
SOURCE            Petrotux sanguineus.
ORGANISM          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
                  Acanthomorpha; Paracanthopterygii; Ophidiiformes; Ophidiidae;
                  Petrotux.
REFERENCE          1 (bases 1 to 218)
AUTHORS            Wiley,E.O., David Johnson,G. and Wheaton Dimmick,W.
TITLE             The interrelationships of Acanthomorph fishes: A total evidence
                  approach using molecular and morphological data
JOURNAL            Biochemical systematics and ecology. 28 (4), 319-350 (2000)
PUBMED            10725591
REFERENCE          2 (bases 1 to 218)
AUTHORS            Wiley,E.O., Johnson,G.D. and Dimmick,W.W.
TITLE             Direct Submission
JOURNAL            Submitted (12-MAY-1999) Natural History Museum, University of
                  Kansas, Lawrence, KS 66045, USA
FEATURES           Location/Qualifiers
                    1..218
                    /organism="Petrotux sanguineus"
                    /specimen_voucher="KU 22948"
                    /db_xref="taxon:94932"
                    <1..>218
                    /product="28S ribosomal RNA"
BASE COUNT        45 a 57 c 86 g 30 t

rRNA
Query Match          100.0%; Score 12; DB 5; Length 218;
Best Local Similarity 100.0%; Pred. No. 5.3e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
|||||
Db 43 CAGGAGCCTCGA 54

REFERENCE          1 (bases 1 to 218)
AUTHORS            Wiley,E.O., David Johnson,G. and Wheaton Dimmick,W.
TITLE             The interrelationships of Acanthomorph fishes: A total evidence
                  approach using molecular and morphological data
JOURNAL            Biochemical systematics and ecology. 28 (4), 319-350 (2000)
PUBMED            10725591
REFERENCE          2 (bases 1 to 218)
AUTHORS            Wiley,E.O., Johnson,G.D. and Dimmick,W.W.
TITLE             Direct Submission
JOURNAL            Submitted (12-MAY-1999) Natural History Museum, University of
                  Kansas, Lawrence, KS 66045, USA
FEATURES           Location/Qualifiers
                    1..218
                    /organism="Petrotux sanguineus"
                    /specimen_voucher="KU 22948"
                    /db_xref="taxon:94932"
                    <1..>218
                    /product="28S ribosomal RNA"
BASE COUNT        45 a 57 c 86 g 30 t

rRNA
Query Match          100.0%; Score 12; DB 5; Length 218;
Best Local Similarity 100.0%; Pred. No. 5.3e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
|||||
Db 43 CAGGAGCCTCGA 54

REFERENCE          1 (bases 1 to 218)
AUTHORS            Wiley,E.O., Johnson,G.D. and Dimmick,W.W.
TITLE             Direct Submission
JOURNAL            Submitted (12-MAY-1999) Natural History Museum, University of
                  Kansas, Lawrence, KS 66045, USA
FEATURES           Location/Qualifiers
                    1..218
                    /organism="Mastacembelus sp."
                    /specimen_voucher="KU 22982"
                    /db_xref="taxon:94235"
                    <1..>218
                    /product="28S ribosomal RNA"
BASE COUNT        45 a 56 c 85 g 32 t

rRNA
Query Match          100.0%; Score 12; DB 5; Length 218;
Best Local Similarity 100.0%; Pred. No. 5.3e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
|||||
Db 43 CAGGAGCCTCGA 54

REFERENCE          1 (bases 1 to 218)
AUTHORS            Wiley,E.O., Johnson,G.D. and Dimmick,W.W.
TITLE             Direct Submission
JOURNAL            Submitted (12-MAY-1999) Natural History Museum, University of
                  Kansas, Lawrence, KS 66045, USA
FEATURES           Location/Qualifiers
                    1..218
                    /organism="Mastacembelus sp."
                    /specimen_voucher="KU 22982"
                    /db_xref="taxon:94235"
                    <1..>218
                    /product="28S ribosomal RNA"
BASE COUNT        45 a 56 c 85 g 32 t

rRNA
Query Match          100.0%; Score 12; DB 8; Length 258;
Best Local Similarity 100.0%; Pred. No. 5.1e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
|||||
Db 99 CAGGAGCCTCGA 88

RESULT 5
G49233            442 bp mRNA linear STS 31-AUG-1999
LOCUS             D1Ertd754e Mouse eight-cell stage embryo cDNA library (M. Ko) Mus
DEFINITION

```

musculus STS cDNA clone J0521C08 3', sequence tagged site.
 G49233
 VERSION
 G49233.1 GI:4726066
 STS
 house mouse.
 SOURCE
 Mus musculus
 ORGANISM
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 442)
 REFERENCE
 Bergstrom,D.E., Gagnon,L.H. and Eicher,E.M.
 Genetic and physical mapping of the dreher locus on mouse
 chromosome 1
 JOURNAL
 Genomics 59 (3), 291-299 (1999)
 MEDLINE
 99375322
 PUBMED
 10444330
 COMMENT

Contact: Eva M. Eicher
 The Eicher Laboratory
 The Jackson Laboratory
 600 Main Street, Bar Harbor, ME 04609 USA
 Tel: 207/288-6344
 Fax: 207/288-6077
 Email: emeejax.org
 Primer A: GCACCTACGGTAATCAAGGC
 Primer B: ACCTGAGTGATGTTGGCT
 STS size: 200
 PCR Profile:

Presoak: 97.0 degrees C for 30 seconds
 Denaturation: 94.0 degrees C for 15 seconds
 Annealing: 55.0 degrees C for 30 seconds
 Polymerization: 72.0 degrees C for 30 seconds + 1 sec/cyc
 Postsoak: 72.0 degrees C for 10 minutes
 PCR Cycles: 40

Protocol:
 Template: 25 ng
 Primer: each 0.25 uM
 dNTPs: each 200 uM
 Taq polymerase: 0.025 units/ul
 Total volume: 20 ul

Buffer:
 MgCl2: 1.5 mM
 KCl: 50 mM
 Tris-HCl: 10 mM
 pH: 8.3

STS designed from Mouse Genome Database DNA segment DI8td754e.

FEATURES
 source
 Location/Qualifiers
 1..442
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="J0521C08"
 /clone_lib="Mouse eight-cell stage embryo cDNA library (M.
 ko)"
 /dev_stage="eight-cell stage embryo"
 23..222
 primer_bind
 23..42
 primer_bind
 complement(203..222)
 BASE COUNT 136 a 90 c 140 g 75 t 1 others
 ORIGIN

Query Match 100.0%; Score 12; DB 11; Length 442;
 Best Local Similarity 100.0%; Pred. No. 4.8e+03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggagctcga 12
 |||||
 Db 374 CAGGAGCTCGA 385

RESULT 6

E12962
 LOCUS
 DEFINITION
 CDNA encoding fragment of phosphoenolpyruvate carboxylase from
 Aloe.
 E12962
 VERSION
 E12962.1 GI:3251791
 JP 1997107975-A/4.
 KEYWORDS
 unidentified.
 SOURCE
 unidentified.
 ORGANISM
 unclassified.
 1 (bases 1 to 450)
 REFERENCE
 Shimada,H., Okamoto,T., Honda,H. and Fujimura,T. .
 PEP-ASE GENE OF MONOCOTYLEDON CAM PLANT
 Patent: JP 1997107975-A 4 28-APR-1997;
 JOURNAL
 MITSUI TOATSU CHEM INC
 COMMENT
 OS Aloe arborescens
 PN JP 1997107975-A/4
 PD 28-APR-1997
 PF 19-OCT-1995 JP 1995294986
 PI SHIMADA HIROAKI, OKAMOTO TOMOKO, HONDA HIDEO, FUJIMURA TATSUTO
 PC C12N15/09,A01H5/00,C07H21/04,C12N1/21,C12N5/10,C12N9/86, PC
 (C12N1/21,
 PC C12R1:19),(C12N5/10,C12R1:91);
 CC strandedness: Double;
 CC topology: Linear;
 CC hypothetical: No;
 CC anti-sense: No;
 FH Key
 FH Location/Qualifiers
 FT source
 1..450
 /organism="Aloe arborescens"
 /tissue_type="green leaf"
 FT mat_peptide
 1..450
 /product="Phosphoenolpyruvate carboxylase from
 Aloe"
 FT /partial.
 FT Location/Qualifiers
 1..450
 /organism="unidentified"
 /db_xref="taxon:32644"

BASE COUNT 96 a 101 c 147 g 106 t
 ORIGIN

Query Match 100.0%; Score 12; DB 6; Length 450;
 Best Local Similarity 100.0%; Pred. No. 4.8e+03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggagctcga 12
 |||||
 Db 424 CAGGAGCTCGA 435

RESULT 7
 HS301F5
 LOCUS
 DEFINITION
 H.sapiens chromosome 22 CpG island DNA genomic MseI fragment, clone
 301f5, complete read.
 ACCESSION
 Z79832
 VERSION
 Z79832.1 GI:1531822
 KEYWORDS
 Chromosome 22; CpG island.
 SOURCE
 human.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 453)
 REFERENCE
 Clark,V.H., Cross,S.H., Simmen,M.W., Langford,C., Carter,N.,
 Bickmore,W. and Bird,A.P.
 Direct Submission
 Submitted (02-SEP-1996) ICMB, University of Edinburgh, King's
 Buildings, Mayfield Rd, Edinburgh EH9 3UR, UK. E-mail contact:
 VCLARK@srvo.bio.ed.ac.uk
 2 (bases 1 to 453)
 JOURNAL
 TITLE

Query Match 100.0%; Score 12; DB 6; Length 450;
 Best Local Similarity 100.0%; Pred. No. 4.8e+03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggagctcga 12
 |||||
 Db 374 CAGGAGCTCGA 385

RESULT 6

Query Match 100.0%; Score 12; DB 9; Length 498;
 Best Local Similarity 100.0%; Pred. No. 4.7e+03; Indels 0; Gaps 0;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
 |||||
 Db 176 CAGGAGCCTCGA 165

RESULT 11
 G39052
 LOCUS Z11764 Zebrafish AB Danio rerio STS genomic, sequence tagged site.
 DEFINITION
 G39052
 ACCESSION
 G39052.1 GI:3358261
 VERSION
 STS.
 zebraphish.
 SOURCE
 ORGANISM
 Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.
 1 (bases 1 to 598)
 Shimoda,N., Knapik,E.W., Ziniti,J., Sim,C., Yamada,E., Kaplan,S.,
 Jackson,D., de Sauvage,F., Jacob,H. and Fishman,M.C.
 Zebrafish genetic map with 2000 microsatellite markers
 Genomics 58 (3), 219-232 (1999)
 93033552
 PUBMED
 10373319

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED

COMMENT
 Contact: Mark C. Fishman
 Cardiovascular Research Center
 Massachusetts General Hospital
 Mail code 1494100A, 149 13th Street, Charlestown, MA 02129, USA
 Fax: 6177263806
 Email: fishman@mgh.cvr.harvard.edu
 http://zebraphish.mgh.harvard.edu
 Primer A: GCGACCTTGTGATGTTT
 Primer B: CATCTGATTGCTTGTCTCA
 STS size: 297
 PCR Profile:
 Presoak: 94 degrees C for 5.0 minutes
 Denaturation: 94 degrees C for 1.0 minute
 Annealing: 58 degrees C for 1.0 minute
 Polymerization: 72 degrees C for 1.5 minute
 PCR Cycles: 27
 Thermal Cycler: MJ Research PTC-100
 Protocol:
 Template: 10 ng
 Primer: each 375 nM
 dNTPs: each 200 uM
 Taq Polymerase: 0.034 units/ul
 Total Vol: 10 ul
 Buffer:
 MgCL2: 1.5 mM
 KCL: 50 mM
 Tris-HCL: 10 mM
 PH: 8.3

Primers are available from Research Genetics Inc.
 (http://www.resgen.com phone: 800-533-4363).

FEATURES
 source
 1..598
 /organism="Danio rerio"
 /strain="AB"
 /db_xref="taxon:7955"
 /sex="F"
 /clone_lib="zebraphish AB"
 /dev_stage="Adult"
 /lab_host="DH5alpha/IQ"
 /note="vector: ml3mp19 with added BstXI site; V-type: Phage; Genomic DNA from a single adult Zebrafish of AB

strain was digested with AluI, Cac8I, HaeIII, NlaVI, or RsaI. Fragments in the range of 250-500 bp were gel purified and a BstXI linker was added. The fragments were cloned into a modified M13mp19 vector and transformed into E. Coli DH5alpha. Microsatellite sequences were screened with labeled d(CA)15 and d(GT)15 oligonucleotide probes."

STS
 primer_bind 123..419
 primer_bind complement(400..419) 155 t 15 others
 BASE COUNT 137 a 149 c 142 g 155 t

ORIGIN
 Query Match 100.0%; Score 12; DB 11; Length 598;
 Best Local Similarity 100.0%; Pred. No. 4.6e+03; Indels 0; Gaps 0;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
 |||||
 Db 82 CAGGAGCCTCGA 93

RESULT 12
 AF321195
 LOCUS Uncultured bacterium rcr_15 680 bp DNA linear BCT 24-JAN-2001
 DEFINITION
 ACCESSION AF321195
 VERSION AF321195.1 GI:12407991
 KEYWORDS
 SOURCE uncultured bacterium rcr_15
 ORGANISM uncultured bacterium rcr_15
 Bacteria; environmental samples.
 REFERENCE 1 (bases 1 to 680)
 AUTHORS Boomer,S.M., Dutton,B.E. and Lodge,D.P.
 TITLE Diversity of novel green non-sulfur bacteria from Yellowstone
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 680)
 AUTHORS Boomer,S.M., Dutton,B.E. and Lodge,D.P.
 TITLE Direct Submission
 JOURNAL Submitted (13-NOV-2000) Dept. of Biology, Western Oregon University, 345 Monmouth Ave., Monmouth, OR 97361, USA
 FEATURES
 source
 1..680
 /organism="uncultured bacterium rcr_15"
 /db_xref="taxon:148410"
 /clone="rcr_15"
 /country="USA: Rabbit Creek, Yellowstone"
 <1..>680
 /product="16S ribosomal RNA"

BASE COUNT 194 a 176 c 192 g 118 t

ORIGIN
 Query Match 100.0%; Score 12; DB 1; Length 680;
 Best Local Similarity 100.0%; Pred. No. 4.5e+03; Indels 0; Gaps 0;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggagcctcga 12
 |||||
 Db 502 CAGGAGCCTCGA 513

RESULT 13
 HUMSYBIAL/c
 LOCUS Human synaptobrevin 1 (SYBI) gene, exon 1.
 DEFINITION
 M36196 J05611
 ACCESSION
 M36196.1 GI:338619
 VERSION
 keywords intrinsic membrane protein; synaptic vesicle protein 2A; synaptobrevin 1.
 SEGMENT 1 of 5

SOURCE Human DNA, clone pcVP18-[1,5].
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 727)
 AUTHORS Archer, B.T. III, Ozcelik, T., Jahn, R., Francke, U. and Sudhof, T.C.
 TITLE Structures and chromosomal localizations of two human genes
 encoding synaptobrevins 1 and 2
 JOURNAL J. Biol. Chem. 265 (28), 17267-17273 (1990)
 MEDLINE 91009161
 COMMENT Draft entry and computer-readable sequence for [J. Biol. Chem.
 (1990) in press] kindly submitted
 by T.C.Sudhof, 06-JUL-1990.
 FEATURES
 source
 1..727
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="12p"
 TATA_signal 544..549
 exon <721..722
 /gene="SVB1"
 /note="synaptobrevin 1 (SVB1); G00-125-292"
 /number=1
 intron 723..>727
 /gene="SVB1"
 /note="synaptobrevin 1 intron A"
 BASE COUNT 130 a 217 c 257 g 123 t
 ORIGIN Chromosome 12.
 Query Match 100.0%; Score 12; DB 9; Length 727;
 Best Local Similarity 100.0%; Pred. No. 4.4e+03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 caggagcctcga 12
 |||||
 Db 84 CAGGAGCCTCGA 73
 RESULT 14
 G41958/c
 LOCUS G41958 761 bp DNA linear STS 30-SEP-1998
 DEFINITION SHGC-68982 Human Homo sapiens STS genomic, sequence tagged site.
 ACCESSION G41958
 VERSION G41958.1 GI:3668291
 KEYWORDS STS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 761)
 AUTHORS Myers, R.M.
 TITLE Human STSs (1998)
 JOURNAL Unpublished
 COMMENT
 Contact: Richard M. Myers
 Stanford Human Genome Center (SHGC)
 Stanford University School of Medicine
 Department of Genetics, M-344, Stanford, CA 94305, USA
 Tel: 4157259687
 Fax: 4157259689
 Email: myers@shgc.stanford.edu
 Primer A: CAGGCAAAAGACTACCACTT
 Primer B: ATGTTCTCTTGTTCTCTGAC
 STS size: 350
 PCR Profile:
 Initial incubation: 95 degrees C for 10 minutes
 Denaturation: 94 degrees C for 30 seconds
 Annealing: 60 degrees C for 30 seconds
 Polymerization: 72 degrees C for 23 seconds
 PCR Cycles: 30
 Thermal Cycler: Perkin Elmer 9700

Protocol:
 Template: 25 ng
 each 1 uM
 Primer: each 200 uM
 dNTPs: each 200 uM
 AmpliTaq Gold Polymerase: 0.07 units/ul
 Total Vol: 5 ul
 Buffer:
 MgCl2: 2.5 mM
 KCl: 50 mM
 Tris-HCl: 10 mM
 pH: 8.3
 Cosmid end sequences from generated from flow-sorted chromosome 4
 (Human Genome Center, Los Alamos National Laboratory), sequenced at
 the Stanford Human Genome Center.
 FEATURES
 source
 1..761
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="4"
 /clone_lib="Human"
 STS 220..569
 primer_bind 220..242
 complement(547..569)
 primer_bind 135 a 235 c 204 g 187 t
 BASE COUNT
 ORIGIN
 Query Match 100.0%; Score 12; DB 11; Length 761;
 Best Local Similarity 100.0%; Pred. No. 4.4e+03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 caggagcctcga 12
 |||||
 Db 298 CAGGAGCCTCGA 287
 RESULT 15
 A70150/c
 LOCUS A70150 822 bp DNA linear PAT 07-MAY-1999
 DEFINITION Sequence 11 from Patent WO9811228.
 ACCESSION A70150
 VERSION A70150.1 GI:4774565
 KEYWORDS
 SOURCE unidentified.
 ORGANISM unidentified.
 REFERENCE 1 (bases 1 to 822)
 AUTHORS Seymour, G.B., Bird, C.R. and Medina-Suarez, R.D.
 TITLE GENETIC CONTROL OF FRUIT RIPENING
 JOURNAL Patent: WO 9811228-A 11 19-MAR-1998;
 SEYMOUR GRAHAM BARRON (GB)
 FEATURES
 source
 1..822
 /organism="unidentified"
 /db_xref="taxon:32644"
 /clone="U-055"
 BASE COUNT 145 a 254 c 224 g 159 t 40 others
 ORIGIN
 Query Match 100.0%; Score 12; DB 6; Length 822;
 Best Local Similarity 100.0%; Pred. No. 4.4e+03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 caggagcctcga 12
 |||||
 Db 310 CAGGAGCCTCGA 299
 Search completed: July 31, 2002, 14:01:55

Thu Aug 1 08:30:03 2002

us-09-899-718a-6.rge

Page 7

Job time: 17557 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 12:21:09 ; Search time 165.21 Seconds
(without alignments)
38.657 Million cell updates/sec

Title: US-09-899-718a-5
Perfect score: 26
Sequence: 1 cccgtctaggcgttcggttcggcc 26

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	18.6	71.5	1248	4	US-09-105-537-7
C 2	18.6	71.5	13613	4	US-09-105-537-3
C 3	16.6	63.8	3061	1	US-09-570-842-1
C 4	16.6	63.8	3061	1	US-08-700-576-1
C 5	16.6	63.8	15664	1	US-08-402-282-3
C 6	16.6	63.8	15664	1	US-08-508-004-3
C 7	16.6	63.8	15664	1	US-08-402-066-3
C 8	16.6	63.8	15664	1	US-08-402-068-3
C 9	16.6	63.8	43280	2	US-08-804-227C-1
C 10	16.4	63.1	4403765	4	US-09-103-840A-2
C 11	16.4	63.1	4411529	4	US-09-103-840A-1
C 12	16	61.5	624	1	US-07-661-610C-14
C 13	16	61.5	1008	4	US-09-199-637A-148
C 14	16	61.5	1392	4	US-08-957-351-1
C 15	16	61.5	2101	2	US-08-637-763B-5
C 16	16	61.5	2101	3	US-09-170-354-5
C 17	16	61.5	5437	1	US-07-661-610C-1
C 18	16	61.5	44377	2	US-08-804-227C-7
C 19	16	61.5	44377	2	US-08-804-198-1
C 20	16	61.5	80161	3	US-09-036-987A-1
C 21	16	61.5	80161	4	US-09-370-700-1
C 22	15.8	60.8	68750	3	US-09-335-409-1
C 23	15.8	60.8	68750	4	US-09-568-102-1
C 24	15.8	60.8	68750	4	US-09-567-969-1
C 25	15.8	60.8	68750	4	US-09-568-480-1
C 26	15.8	60.8	68750	4	US-09-568-486-1
C 27	15.8	60.8	68750	4	US-09-568-472-1

```

c 28 15.8 60.8 71989 4 US-09-443-501A-2 Sequence 2, Appli
c 29 15.6 60.0 1941 4 US-09-232-200-28 Sequence 28, Appl
c 30 15.6 60.0 1941 4 US-09-232-197-28 Sequence 28, Appl
c 31 15.6 60.0 1941 4 US-09-232-201-28 Sequence 28, Appl
c 32 15.6 60.0 2220 4 US-08-765-907A-14 Sequence 14, Appl
c 33 15.6 60.0 3694 4 US-09-232-197-46 Sequence 46, Appl
c 34 15.6 60.0 3694 4 US-09-232-197-46 Sequence 46, Appl
c 35 15.6 60.0 3694 4 US-09-232-201-46 Sequence 46, Appl
c 36 15.6 60.0 3704 4 US-09-232-200-24 Sequence 24, Appl
c 37 15.6 60.0 3704 4 US-09-232-197-24 Sequence 24, Appl
c 38 15.6 60.0 3704 4 US-09-232-201-24 Sequence 24, Appl
c 39 15.6 60.0 4496 4 US-08-765-907A-6 Sequence 6, Appl
c 40 15.6 60.0 18318 1 US-08-414-926A-6 Sequence 6, Appl
c 41 15.6 60.0 18318 2 US-08-926-922-6 Sequence 6, Appl
c 42 15.6 60.0 18318 3 US-09-253-682-6 Sequence 6, Appl
c 43 15.6 60.0 18318 4 US-09-527-657-6 Sequence 6, Appl
c 44 15.4 59.2 960 1 US-07-639-330E-1 Sequence 1, Appli
c 45 15.4 59.2 960 3 US-09-248-335-57 Sequence 57, Appl

```

ALIGNMENTS

```

RESULT 1
US-09-105-537-7/c
; Sequence 7, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: fastseq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 1248
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-7

```

```

Query Match 71.5%; Score 18.6; DB 4; Length 1248;
Best Local Similarity 84.0%; Pred. No. 5;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

Qy 1 cccgtctaggcgttcggttcggcc 25
Db 1068 CGCGGCGAGCGGTCGGTGTGCGGC 1044

```

```

RESULT 2
US-09-105-537-3
; Sequence 3, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 13613
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae

```



```
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 747..1109
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1109..2014
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2034..2747
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2747..3109
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3109..3444
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3444..3728
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3731..4855
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4855..5376
OTHER INFORMATION: /function= "potential coding
OTHER INFORMATION: sequence"
OTHER INFORMATION: /product= "L5 gp37 homolog"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 5382..5747
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 5837..6307
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6403..7770
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7770..8006
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 8033..8236
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 8244..9443
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
```

```
NAME/KEY: misc_feature
LOCATION: 9450..10244
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10371..10586
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11115..11786
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11917..12741
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 12748..14499
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14771..15154
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15154..15426
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15429..15664
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
US-08-402-282-3
```

Query Match 63.8%; Score 16.6; DB 1; Length 15664;
Best Local Similarity 82.6%; Pred. No. 44;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 cgtctaggcgttcggtccggc 25
DB 11139 CGTCCAGGATTCGGTGTCGTC 11117

RESULT 6
US-08-508-004-3/c
Sequence 3, Application US/08508004
Patent No. 5582969
GENERAL INFORMATION:
APPLICANT: Pearson, Robert E.
APPLICANT: Dickson, Julie A.
APPLICANT: Hamilton, Paul T.
APPLICANT: Little, Michael C.
APPLICANT: Beyer Jr., Wayne F.
TITLE OF INVENTION: MYCOBACTERIOPHAGE SPECIFIC FOR THE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Richard J. Rodrick, Becton Dickinson and
ADDRESSEE: Company
STREET: 1 Becton Drive
CITY: Franklin Lakes
STATE: NJ
COUNTRY: US
ZIP: 07417
COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/508,004
FILING DATE: 27-JUL-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/402,282
FILING DATE: 10-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Fugit, Donna R.
REGISTRATION NUMBER: 32,135
REFERENCE/DOCKET NUMBER: P-3293
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 15664 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: misc_feature
LOCATION: 222..425
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 451..747
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 747..1109
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1109..2014
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2034..2747
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2747..3109
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3109..3444
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3444..3728
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3731..4855
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4855..5376
OTHER INFORMATION: /function= "potential coding
OTHER INFORMATION: sequence"
OTHER INFORMATION: /product= "L5 qp37 homolog"

```

Query Match 63.8%; Score 16.6; DB 1; Length 15664;
Best Local Similarity 82.6%; Pred. No. 44;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 cgtctaggcgttcggttcgcgcg 25
||||| ||| ||||| ||||| |||
Db 11139 GGTCCAGGATTGCGTGTCGTC 11117

RESULT 7

US-08-402-066-3/c

; Sequence 3, Application US/08402066

; Patent No. 5612182

; GENERAL INFORMATION:

; APPLICANT: Pearson, Robert E.

; APPLICANT: Dickson, Julie A.

; APPLICANT: Hamilton, Paul F.

; APPLICANT: Little, Michael C.

; APPLICANT: Beyer Jr., Wayne F.

; TITLE OF INVENTION: MYCOBACTERIOPHAGE SPECIFIC FOR THE

; TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS COMPLEX

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Richard J. Rodrick, Becton Dickinson and

; ADDRESSEE: Company

; STREET: 1 Becton Drive

; CITY: Franklin Lakes

; STATE: NJ

; COUNTRY: US

; ZIP: 07417

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/402,066

; FILING DATE:

; CLASSIFICATION: 436

; ATTORNEY/AGENT INFORMATION:

; NAME: Fugit, Donna R.

; REGISTRATION NUMBER: 32,135

; REFERENCE/DOCKET NUMBER: P-3283

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 15664 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: 222..425

; OTHER INFORMATION: /function= "potential open reading

; OTHER INFORMATION: frame"

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: 451..747

; OTHER INFORMATION: /function= "potential open reading

; OTHER INFORMATION: frame"

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: 747..1109

; OTHER INFORMATION: /function= "potential open reading

; OTHER INFORMATION: frame"

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: 1109..2014

; OTHER INFORMATION: /function= "potential open reading

; OTHER INFORMATION: frame"

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: 1115..11786

; OTHER INFORMATION: /function= "potential open reading

; OTHER INFORMATION: frame"

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: 2034..2747

OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2747..3109
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3109..3444
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3444..3728
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3731..4855
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4855..5376
OTHER INFORMATION: /function= "potential coding
OTHER INFORMATION: sequence"
OTHER INFORMATION: /product= "L5 gp37 homolog"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 5382..5747
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 5837..6307
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6403..7770
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7770..8006
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 8033..8236
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 8244..9443
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 9450..10244
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10371..10586
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11115..11786
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"

```
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11917..12741
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 12748..14499
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14771..15154
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15154..15426
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15429..15664
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
US-08-402-066-3

Query Match 63.8%; Score 16.6; DB 1; Length 15664;
Best Local Similarity 82.6%; Pred. No. 44;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 cgtctaggcgttcgtgtccgc 25
    ||||| ||| ||||| ||||| ||
Db 11139 CGTCCAGGATTGGTGTCGTC 11117

RESULT 8
US-08-402-068-3/c
Sequence 3, Application US/08402068
Patent No. 5633159
GENERAL INFORMATION:
APPLICANT: Pearson, Robert E.
APPLICANT: Dickson, Julie A.
APPLICANT: Hamilton, Paul T.
APPLICANT: Little, Michael C.
APPLICANT: Beyer Jr., Wayne F.
TITLE OF INVENTION: MYCOBACTERIOPHAGE SPECIFIC FOR THE
TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS COMPLEX
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Richard J. Rodrick, Becton Dickinson and
Company
STREET: 1 Becton Drive
CITY: Franklin Lakes
STATE: NJ
COUNTRY: US
ZIP: 07417
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/402,068
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Fugit, Donna R.
REGISTRATION NUMBER: 32,135
REFERENCE/DOCKET NUMBER: P-3283
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 15664 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: misc_feature
LOCATION: 222..425
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 451..747
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 747..1109
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1109..2014
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2034..2747
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2747..3109
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3109..3444
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3444..3728
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3731..4855
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4855..5376
OTHER INFORMATION: /function= "potential coding
sequence"
OTHER INFORMATION: /product= "L5 gp37 homolog"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 5382..5747
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 5837..6307
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6403..7770
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
```



```
; LOCATION: 7770..8006
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 8033..8236
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 8244..9443
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 9450..10244
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10371..10586
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11115..11786
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11917..12741
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 12748..14499
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 14771..15154
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 15154..15426
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 15429..15664
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; US-08-402-068-3
```

```
Query Match 63.8%; Score 16.6; DB 1; Length 15664;
Best Local Similarity 82.6%; Pred. No. 44;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy 3 cgtctaggcgttcggtgctccggc 25
    ||| ||| ||| ||| ||| ||| |||
Db 11139 CGTCAGGATTTCGGTGTCGTC 11117
```

```
RESULT 9
US-08-402-227C-1/c
; Sequence 1, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rosteck, Paul R., Jr.
```

```
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCI(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 43280 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 816..14234
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14351..19945
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20010..31199
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31232..36067
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36249..41774
; US-08-804-227C-1
```

```
Query Match 63.8%; Score 16.6; DB 2; Length 43280;
Best Local Similarity 82.6%; Pred. No. 46;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy 3 cgtctaggcgttcggtgctccggc 25
    ||| ||| ||| ||| ||| ||| |||
Db 27009 CGTCAGCGCTCGGTGTCGAGC 26987
```

```
RESULT 10
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
```

```

RESULT 12
US-07-661-610C-14/c
; Sequence 14, Application US/07661610C
; Patent No. 5292643
; GENERAL INFORMATION:
; APPLICANT: Shibano, Yuji
; APPLICANT: Toyoda, Hideyoshi
; APPLICANT: Utsumi, Ryutaro
; APPLICANT: Obata, Kazuaki
; TITLE OF INVENTION: Fusaric Acid Resistant Genes
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLOK, SPINAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; ZIP: 22202

```

Qy 3 cgtctaggcggttcggtgtccggcc 26
||| ||| ||| ||| ||| ||| ||| |||
Db 475 CGTTCAGGCCCTTCGGTGACCCAGCC 452

RESULT 14
US-08-957-351-1/c
; Sequence 1, Application US/08957351
; Patent No. 6306586
; GENERAL INFORMATION:
; APPLICANT: Semina, Elena
; APPLICANT: Murray, Jeffrey C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF CATARACTS
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/957,351
; FILING DATE: 24-OCT-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold, Beth E.
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: UTA-024.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1392 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-08-957-351-1

Query Match 61.5%; Score 16; DB 4; Length 1392;
Best Local Similarity 79.2%; Pred. No. 74;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 cccgtctaggcgttcggtgtccgg 24
|| ||||| |||||
Db 1119 CACGTCTAGCGACCCCTGTCCGG 1096

RESULT 15
US-08-637-763B-5
; Sequence 5, Application US/08637763B
; Patent No. 5849559
; GENERAL INFORMATION:
; APPLICANT: VAN DER WOUW, Monique J.A. et al
; TITLE OF INVENTION: ARABINOXYLAN DEGRADING ENZYME
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/637,763B
; FILING DATE: 25-AUG-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murashige, Kate H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 4615-0066.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030 MRSNFOERSWSH
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2101 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus niger
; STRAIN: CBS 120.49
; FEATURE:
; NAME/KEY: TATA_signal
; LOCATION: 665...670
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 784...1779
; OTHER INFORMATION: /product= "arabinoxylan degrading
; OTHER INFORMATION: enzyme" /gene= "axda"
; OTHER INFORMATION: /standard_name= "arabinoxylan degrading enzyme"
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 784...861
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 862...1779
US-08-637-763B-5

Query Match 61.5%; Score 16; DB 2; Length 2101;
Best Local Similarity 79.2%; Pred. No. 75;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 cccgtctaggcgttcggtgtccgg 24
| ||||| | |||||
Db 266 CACGTCTAGACTTCGATGCCCGG 289

Search completed: July 31, 2002, 12:23:47
Job time: 11850 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 12:12:01 ; Search time 5855.71 Seconds
(without alignments)
59.928 Million cell updates/sec

Title: US-09-899-718A-5
Perfect score: 26
Sequence: 1 cccgtctaggcgttcggtgcgcgc 26

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_estli:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pin:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	19.2	73.8	668	12	BH506552	BH506552 BOGFR94TF
2	19.2	73.8	747	12	BH428842	BH428842 BOGFR94TF
3	18.6	71.5	420	12	BH391002	BH391002 AG-ND-180
4	18.6	71.5	700	10	BM072919	BM072919 MEST55-H0
5	18.2	70.0	576	10	BF770774	BF770774 RCI-IT001
6	18.2	70.0	771	12	CNS02PLG	AL208141 Tetraodon
7	18.6	69.2	319	12	A2496643	A2496643 IM0333K08
8	18.6	69.2	745	10	BM398961	BM398961 5009-0-51
9	18.6	69.2	845	10	BE748192	BE748192 601571585
10	18.6	69.2	1031	12	AG060278	AG060278 Pan trogl
11	18.6	69.2	1425	10	B8669446	BE8669446 601446101
12	17.8	68.5	633	9	AW949909	AW949909 EST361979
13	17.6	67.7	495	10	BG818796	BG818796 602779143
14	17.6	67.7	499	10	BE462300	BE462300 EST324564
15	17.6	67.7	591	10	BI490136	BI490136 603031972
16	17.6	67.7	652	10	BI925736	BI925736 EST545625
17	17.6	67.7	702	10	BF867684	BF867684 963093D05

C 18	17.6	67.7	824	10	BE892849	BE892849 601433981
C 19	17.6	67.7	867	10	BF241175	BF241175 601880044
C 20	17.6	67.7	877	10	BI462741	BI462741 603202191
C 21	17.6	67.7	911	10	BI601214	BI601214 603245726
C 22	17.6	67.7	932	10	BF180488	BF180488 601805280
C 23	17.6	67.7	1147	10	BI834829	BI834829 603090281
C 24	17.2	66.2	215	10	BF770115	BF770115 RCI-IT001
C 25	17.2	66.2	223	10	BE551977	BE551977 hy02b02.x
C 26	17.2	66.2	314	9	AL133850	AL133850 DKFp761F
C 27	17.2	66.2	348	10	BE937644	BE937644 MRI-TN002
C 28	17.2	66.2	348	10	BE937659	BE937659 MRI-TN002
C 29	17.2	66.2	355	10	BE937624	BE937624 MRI-TN002
C 30	17.2	66.2	376	10	N60925	N60925 TgESTzyl4d0
C 31	17.2	66.2	421	10	H20321	H20321 ym62q01.s1
C 32	17.2	66.2	438	10	H15755	H15755 y127d10.r1
C 33	17.2	66.2	458	9	AI222008	AI222008 qg95d11.x
C 34	17.2	66.2	520	9	BE161596	BE161596 MRO-HT044
C 35	17.2	66.2	539	10	BI774972	BI774972 467308 MA
C 36	17.2	66.2	603	10	BG992087	BG992087 MR2-HT116
C 37	17.2	66.2	607	10	BG982963	BG982963 PMO-CN015
C 38	17.2	66.2	610	9	AI398765	AI398765 NCW09A777
C 39	17.2	66.2	665	10	BG983477	BG983477 PMO-CN015
C 40	17.2	66.2	681	10	BG419972	BG419972 602453646
C 41	17.2	66.2	707	10	BG479901	BG479901 602705750
C 42	17.2	66.2	733	10	BG428650	BG428650 602494788
C 43	17.2	66.2	756	10	BE727948	BE727948 601560594
C 44	17.2	66.2	818	10	BF316519	BF316519 601903032
C 45	17.2	66.2	847	10	BG281333	BG281333 602401814

ALIGNMENTS

RESULT 1
BH506552
LOCUS
DEFINITION
BOGFR94TF BOGF Brassica oleracea genomic clone BOGFR94, DNA
sequence.
ACCESSION
BH506552
VERSION
BH506552.1 GI:17714649
KEYWORDS
GSS.
SOURCE
Brassica oleracea.
ORGANISM
Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE
1 (bases 1 to 668)
Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished (2001)
Other GSSs: BOGFR94TF
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
Location/Qualifiers
1.668
/organism="Brassica oleracea"
/strain="TCL000DH3"
/db_xref="taxon:3712"
/clone="BOGFR94"
/clone_lib="BOGF"
/note="Vector: pHS1; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into pHS1 using BstXI linkers"

FEATURES

source
195 a 140 c 125 g 208 t
BASE COUNT
ORIGIN

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 14:11:17 ; Search time 720.8 Seconds

(without alignments)
61.931 Million cell updates/sec

Title: US-09-899-718A-5

Perfect score: 26

Sequence: 1 ccgcctagcgttcggtccgccc 26

Scoring table: IDENTIFY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_032802.*

- 1: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1980.DAT.*
- 2: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1981.DAT.*
- 3: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1982.DAT.*
- 4: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1983.DAT.*
- 5: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1984.DAT.*
- 6: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1985.DAT.*
- 7: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1986.DAT.*
- 8: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1987.DAT.*
- 9: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1988.DAT.*
- 10: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1989.DAT.*
- 11: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1990.DAT.*
- 12: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1991.DAT.*
- 13: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1992.DAT.*
- 14: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1993.DAT.*
- 15: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1994.DAT.*
- 16: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1995.DAT.*
- 17: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1996.DAT.*
- 18: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1997.DAT.*
- 19: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1998.DAT.*
- 20: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA1999.DAT.*
- 21: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA2000.DAT.*
- 22: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA2001A.DAT.*
- 23: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA2001B.DAT.*
- 24: /SIDSL/gcgdata/hold-geneq/geneq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	18.6	71.5	1248	21	AAZ87286
c 2	18.6	71.5	12441	21	AAZ87284
c 3	18.6	71.5	13613	21	AAZ87319
c 4	17.6	67.7	47981	22	AAZ87357
c 5	17.2	66.2	1680	22	AAH41415
c 6	17	65.4	381	22	ABA46067
c 7	17	65.4	381	22	ABA56605
c 8	17	65.4	381	22	ABA26220
c 9	17	65.4	381	22	AAK04749

c 10	17	65.4	381	22	AAK30275	Human bone marrow
c 11	17	65.4	381	22	AAI14871	Probe #4804 for ge
c 12	17	65.4	381	22	AAI36227	Probe #4913 used t
c 13	17	65.4	381	22	AAI04658	Probe #4649 used t
c 14	17	65.4	1463	11	AAQ06685	Polygisticronic oper
c 15	17	65.4	2443	22	AAH35021	Human colon cancer
c 16	17	65.4	32249	22	AAH04931	Human reproductive
c 17	16.6	63.8	120	21	AAI12097	Human secreted pro
c 18	16.6	63.8	3061	19	AAV22799	Nucleotide sequenc
c 19	16.6	63.8	4599	20	AAV57803	G. oxydans autonom
c 20	16.6	63.8	15664	17	AAI09312	Mycobacteriophage
c 21	16.6	63.8	15664	18	AAI70492	Mycobacteriophage
c 22	16.6	63.8	15664	18	AAI66130	Mycobacteriophage
c 23	16.6	63.8	15664	18	AAI51224	NheI-G/SpeI fragme
c 24	16.6	63.8	43280	18	AAI80413	Tylosine synthase
c 25	16.4	63.1	952	23	ABL27597	Drosophila melanog
c 26	16.4	63.1	1175	21	AAI22378	Human secreted pro
c 27	16.4	63.1	1343	22	ABA16862	Human nervous syst
c 28	16.4	63.1	1343	22	ABA19759	Human nervous syst
c 29	16.4	63.1	1571	23	ABL06699	Drosophila melanog
c 30	16.4	63.1	1959	23	ABL18703	Drosophila melanog
c 31	16.4	63.1	2310	22	AAI03788	Mycobacterium tube
c 32	16.4	63.1	2495	12	AAQ11271	Prepro-polygalactu
c 33	16.4	63.1	2547	22	AAI77899	Human cDNA encodin
c 34	16.4	63.1	3090	22	AAI77899	Quorum sensing con
c 35	16.4	63.1	3442	23	ABL07609	Drosophila melanog
c 36	16.4	63.1	3728	23	ABL06698	Drosophila melanog
c 37	16.4	63.1	4109	23	ABL18702	Drosophila melanog
c 38	16.4	63.1	12713	23	ABL07608	Drosophila melanog
c 39	16.4	63.1	24221	23	ABL09808	Drosophila melanog
c 40	16.4	63.1	4403765	22	AAI99683	Mycobacterium tube
c 41	16.4	63.1	4411529	22	AAI99682	Mycobacterium tube
c 42	16.2	62.3	303	22	AAK8761	Human digestive sy
c 43	16.2	62.3	303	22	AAI31795	Human liver associ
c 44	16.2	62.3	528	22	AAI30761	Megalomicin polyke
c 45	16.2	62.3	580	23	ABL25667	Drosophila melanog

ALIGNMENTS

RESULT	1
ID	AAZ87286/C
XX	AAZ87286 standard; DNA; 1248 BP.
AC	AAZ87286;
XX	05-JUN-2000 (first entry)
DT	S. venezuelae desosamine biosynthetic gene desI, SEQ ID NO:7.
DE	Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
DE	neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW	biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma,
KW	chronic obstructive pulmonary disease; respiratory inflammation;
KW	hypercholesterolaemia; crop protection agent; ds.
XX	Streptomyces venezuelae ATCC15439.
OS	Key
XX	Location/Qualifiers
FH	1..1248
FT	FT
FT	/*tag= a
FT	/product= "DesI"
XX	WO200000620-A2.
PN	06-JAN-2000.
PD	25-JUN-1999; 99WO-US14398.
XX	26-JUN-1998; 98US-0105537.
XX	(MINU) UNIV MINNESOTA.

```

XX Sherman DH, Liu H,  Xue Y,  Zhao L;
PI WPI: 2000-160679/14.
DR P-PSDB; AAY77181.
XX
XX Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT synthesis of methymycin and pikromycin -
XX
XX Claim 3; Page 353-354; 438pp; English.
XX
XX The invention relates to an isolated and purified nucleic acid segment
CC comprising a desosamine biosynthetic gene cluster, a fragment or its
CC biologically active variant, where the nucleic acid sequence is not
CC derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC Streptomyces antibioticus. The invention also relates to a macrolide
CC biosynthetic gene cluster, or fragments thereof. The macrolide
CC biosynthetic gene cluster encodes proteins which synthesize methymycin,
CC pikromycin, neomethymycin, narbomycin or a combination of these
CC compounds. Recombinant or augmented cells comprising the desosamine
CC and/or macrolide biosynthetic gene clusters are useful for the
CC production of biologically active macrolides. The macrolide biosynthetic
CC proteins are useful for synthesis of methymycin, pikromycin,
CC neomethymycin and narbomycin. The alternative termination of polyketide
CC synthesis may be useful to prepare novel antibiotics and
CC polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
CC recombinant host cells are useful as biopolymers, e.g., in packaging or
CC biomedical applications, to engineer PHA monomer synthases or to prepare
CC biologically active agents, such as chemotherapeutics,
CC immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
CC disease as well as other diseases involving respiratory inflammation,
CC cholesterol-lowering agents or macrolide-based antibiotics which are
CC active against a variety of organisms, e.g., bacteria, including
CC multi-drug resistant pneumococci and other respiratory pathogens, as well
CC as viral parasitic pathogens, or as crop protection agents (e.g.,
CC fungicides or insecticides) via expression of polyketides in plants.
CC Sequences AA87286-287294 represent desosamine biosynthetic genes from
CC Streptomyces venezuelae ATCC 15439, which encode proteins
XX
XX AAY77181-Y77189.
XX
XX Sequence 1248 BP; 177 A; 546 C; 377 G; 148 T; 0 other;

Query Match      71.5%; Score 18.6; DB 21; Length 1248;
Best Local Similarity 84.0%; Pred. NO. 29;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 cccgtctaggcgttcggtgtccggc 25
Db 1068 CGCGCGAGCGGCTTCGGTGTGCGGC 1044

RESULT 2
AAZ87284
ID AAZ87284 standard; DNA; 12441 BP.
XX
AC AAZ87284;
XX
XX 05-JUN-2000 (first entry)
XX
XX S. venezuelae desosamine biosynthetic gene cluster pikB, SEQ ID NO:3.
XX
XX Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
XX neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
XX biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma,
XX chronic obstructive pulmonary disease; respiratory inflammation;
XX hypercholesterolaemia; crop protection agent; ds.
XX
XX Streptomyces venezuelae ATCC15439.
XX
XX W0200000620-A2.
XX
XX 06-JAN-2000.
XX

```

```

XX 25-JUN-1999; 99WO-US14398.
XX
XX 26-JUN-1998; 98US-0105537.
XX
XX (MINU ) UNIV MINNESOTA.
XX
XX Sherman DH, Liu H,  Xue Y,  Zhao L;
PI WPI: 2000-160679/14.
DR P-PSDB; AAY77179.
XX
XX Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT synthesis of methymycin and pikromycin -
XX
XX Claim 2; Page 281-287; 438pp; English.
XX
XX The invention relates to an isolated and purified nucleic acid segment
CC comprising a desosamine biosynthetic gene cluster, a fragment or its
CC biologically active variant, where the nucleic acid sequence is not
CC derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC Streptomyces antibioticus. The invention also relates to a macrolide
CC biosynthetic gene cluster, or fragments thereof. The macrolide
CC biosynthetic gene cluster encodes proteins which synthesize methymycin,
CC pikromycin, neomethymycin, narbomycin or a combination of these
CC compounds. Recombinant or augmented cells comprising the desosamine
CC and/or macrolide biosynthetic gene clusters are useful for the production
CC of biologically active macrolides. The macrolide biosynthetic proteins
CC are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC narbomycin. The alternative termination of polyketide synthesis may be
CC useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC monomers. The compounds produced by the recombinant host cells are useful
CC as biopolymers, e.g., in packaging or biomedical applications, to
CC engineer PHA monomer synthases or to prepare biologically active agents,
CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC chronic obstructive pulmonary disease as well as other diseases involving
CC respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC antibiotics which are active against a variety of organisms, e.g.,
CC respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC protection agents (e.g., fungicides or insecticides) via expression of
CC polyketides in plants. The present sequence represents the desosamine
CC biosynthetic gene cluster from Streptomyces venezuelae ATCC 15439.
XX
XX Sequence 12441 BP; 1704 A; 4294 C; 4686 G; 1757 T; 0 other;

Query Match      71.5%; Score 18.6; DB 21; Length 12441;
Best Local Similarity 84.0%; Pred. NO. 31;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 cccgtctaggcgttcggtgtccggc 25
Db 11129 cgcgcgcgcgcgttcggtgtccggc 11153

RESULT 3
AAZ87319
ID AAZ87319 standard; DNA; 13613 BP.
XX
AC AAZ87319;
XX
XX 05-JUN-2000 (first entry)
XX
XX S. venezuelae desosamine biosynthetic gene cluster pikB.
XX
XX Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
XX neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
XX biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma,
XX chronic obstructive pulmonary disease; respiratory inflammation;
XX hypercholesterolaemia; crop protection agent; ds.
XX
XX Streptomyces venezuelae ATCC15439.
XX

```

```
XX Key Location/Qualifiers
FH CDS 1..809
FT /tag= a
FT /partial
FT /product= "PikB gene cluster protein #1 (AAV77204)"
FT /note= "No initiation codon given in the specification"
FT CDS 806..2014
FT /tag= b
FT /product= "PikB gene cluster protein #2 (AAV80998)"
FT 2162..6741
FT /tag= c
FT /product= "PikB gene cluster protein #3 (AAV77205)"
FT 6834..7402
FT /tag= d
FT /product= "PikB gene cluster protein #4 (AAV77206)"
FT /transl_except= (pos:6837..6841, aa:Gln)
FT CDS 7492..8205
FT /tag= e
FT /product= "PikB gene cluster protein #5 (AAV77207)"
FT complement (7942..8205)
FT /partial
FT /tag= f
FT /product= "PikB gene cluster protein #6 (AAV77208)"
FT /note= "No termination codon given in the specification"
FT /transl_except= (pos:8270..8272, aa:Val)
FT /transl_except= (pos:8273..8275, aa:Thr)
FT /transl_except= (pos:8276..8278, aa:Gly)
FT CDS complement (10126..11139)
FT /tag= g
FT /product= "PikB gene cluster protein #7 (AAV80999)"
FT complement (11271..12149)
FT /tag= h
FT /product= "PikB gene cluster protein #8 (AAV77209)"
FT complement (12342..13799)
FT /tag= i
FT /product= "PikB gene cluster protein #9 (AAV77210)"
FT complement (13706..15043)
FT /tag= j
FT /product= "PikB gene cluster protein #10 (AAV77211)"
FT 13404..15574
FT /tag= k
FT /product= "PikB gene cluster protein #11 (AAV77212)"
XX WO200000620-A2.
XX 06-JAN-2000.
XX 25-JUN-1999; 99WO-US14398.
XX 26-JUN-1998; 98US-0105537.
XX (MINU ) UNIV MINNESOTA.
XX Sherman DH, Liu H, Xue Y, Zhao L;
XX WPI: 2000-160679/14.
XX P-PSDB: AAV77204, AAV77205, AAV77206, AAV77207, AAV77208.
XX AAV77209, AAV77210, AAV77211, AAV77212, AAV80998, AAV80999.
XX Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.
XX synthesis of methymycin and pikromycin.
XX Disclosure; Figure 32; 438pp; English.
XX The invention relates to an isolated and purified nucleic acid segment
XX comprising a desosamine biosynthetic gene cluster, a fragment or its
XX biologically active variant, where the nucleic acid sequence is not
XX derived from the eryC gene cluster of Saccharopolyspora erythraea or
XX Streptomyces antibioticus. The invention also relates to a macrolide
XX biosynthetic gene cluster, or fragments thereof. The macrolide
XX biosynthetic gene cluster encodes proteins which synthesise methymycin,
XX pikromycin, neomethymycin, narbomycin or a combination of these
```

```
CC compounds. Recombinant or augmented cells comprising the desosamine
CC and/or macrolide biosynthetic gene clusters are useful for the production
CC of biologically active macrolides. The macrolide biosynthetic proteins
CC are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC narbomycin. The alternative termination of polyketide synthesis may be
CC useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC monomers. The compounds produced by the recombinant host cells are useful
CC as biopolymers, e.g., in packaging or biomedical applications, to
CC engineer PHA monomer synthases or to prepare biologically active agents,
CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC chronic obstructive pulmonary disease as well as other diseases involving
CC respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC antibiotics which are active against a variety of organisms, e.g.,
CC bacteria, including multi-drug resistant pneumococci and other
CC respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC protection agents (e.g., fungicides or insecticides) via expression of
CC polyketides in plants. The present sequence represents the desosamine
CC biosynthetic gene cluster (pik) from Streptomyces venezuelae ATCC
CC 15439, as given in figure 32.
XX SQ Sequence 13613 BP; 1858 A; 4733 C; 5092 G; 1930 T; 0 other;

Query Match 71.5%; Score 18.6; DB 21; Length 13613;
Best Local Similarity 84.0%; Pred. No. 31;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 cccgtctaggcggttcggtgcggc 25
Db 11816 cgcgcgcgagcggttcggtgcggc 11840

RESULT 4
AAF30757/C
ID AAF30757 standard; DNA; 47981 BP.
XX AC AAF30757;
XX DT 21-JUN-2001 (first entry)
XX DE Micromonospora megalomicea megalomicin biosynthetic gene cluster.
XX KW Megalomicin; meg gene; polyketide synthase; antibiotic;
XX motilide; antiparasitic; ds.
XX OS Micromonospora megalomicea subsp. nigra.
XX FH Key Location/Qualifiers
FT CDS complement (1..144)
FT /tag= a
FT /partial
FT /gene= "megBVI(megT)"
FT /product= "TDP-4-keto-6-deoxyglucose-
FT 2,3-dehydratase"
FT /note= "encodes AAB82201"
FT 928..2061
FT /tag= b
FT /gene= "megDVI"
FT /product= "TDP-4-keto-6-deoxyhexose 3,4-isomerase"
FT /note= "encodes AAB82202"
FT 2072..3382
FT /tag= c
FT /gene= "megDI"
FT /product= "TDP-megosamine glycosyltransferase"
FT /note= "eryCIII homologue; encodes AAB802203"
FT 3462..4634
FT /tag= d
FT /gene= "megY"
FT /product= "mycarose O-acyltransferase"
FT /note= "encodes AAB82204"
FT 4651..5775
FT /tag= e
FT /gene= "megDII"
FT CDS
```

```
FT FT /product= "TDP-3-keto-6-deoxyhexose
FT FT 3-aminotransaminase"
FT FT /note= "eryCI, DnrJ homologue, encodes AAB82205"
FT FT 5822..6595
FT FT /*tag= f
FT FT /gene= "megDIII"
FT FT /product= "daunosaminyl-N,N-dimethyltransferase"
FT FT /note= "eryCVI homologue; encodes AAB82206"
FT FT 6592..7197
FT FT /*tag= g
FT FT /gene= "megDIV"
FT FT /product= "TDP-4-keto-6-deoxyhexose 3,5-epimerase"
FT FT /note= "eryVIII, dnmV homologue; encodes AAB82207"
FT FT 7220..8206
FT FT /*tag= h
FT FT /gene= "megDV"
FT FT /product= "TDP-4-keto-6-deoxyhexose 4-ketoreductase"
FT FT /note= "eryVIV, dnmV homologue; encodes AAB82208"
FT FT complement (8228..9220)
FT FT /*tag= i
FT FT /gene= "megBII-1(megDVII)"
FT FT /product= "TDP-4-keto-6-deoxyhexose 2,3-reductase"
FT FT /note= "encodes AAB82209"
FT FT complement (9226..10479)
FT FT /*tag= j
FT FT /gene= "megBV"
FT FT /product= "TDP-mycarose glycosyltransferase"
FT FT /note= "encodes AAB82210"
FT FT complement (10483..11424)
FT FT /*tag= k
FT FT /gene= "megBIV"
FT FT /product= "TDP-4-keto-6-deoxyhexose 4-ketoreductase"
FT FT /note= "encodes AAB82211"
FT FT 12181..22821
FT FT /gene= "megA"
FT FT /*tag= l
FT FT /product= "megalomicin 6-deoxyerythronolide B
FT FT synthase 1"
FT FT /note= "polyketide synthase; encodes AAB82212"
FT FT 12505..13470
FT FT /*tag= m
FT FT /gene= "megA"
FT FT /function= "AT-L"
FT FT 13576..13791
FT FT /*tag= n
FT FT /gene= "megA"
FT FT /function= "ACP-L"
FT FT 13849..15126
FT FT /*tag= o
FT FT /gene= "megA"
FT FT /function= "KS1"
FT FT 15427..16476
FT FT /*tag= p
FT FT /gene= "megA"
FT FT /function= "ATl"
FT FT 17155..17694
FT FT /*tag= q
FT FT /gene= "megA"
FT FT /function= "KR1"
FT FT 17947..18207
FT FT /*tag= r
FT FT /gene= "megA"
FT FT /function= "ACP1"
FT FT 18268..19548
FT FT /*tag= s
FT FT /gene= "megA"
FT FT /function= "KS2"
FT FT 19876..20910
FT FT /*tag= t
FT FT /gene= "megA"
FT FT /function= "AT2"
FT FT 21517..22053
FT FT /*tag= u
FT FT /gene= "megA"
FT FT /function= "KR2"
FT FT 22318..22575
FT FT /*tag= v
FT FT /gene= "megA"
FT FT /function= "ACP2"
FT FT 22867..33555
FT FT /*tag= w
FT FT /gene= "megAII"
FT FT /product= "megalomicin 6-deoxyerythronolide B
FT FT synthase 2"
FT FT /note= "polyketide synthase, encodes AAB82213"
FT FT 22957..24237
FT FT /*tag= x
FT FT /gene= "megAII"
FT FT /function= "KS3"
FT FT 24544..25581
FT FT /*tag= y
FT FT /gene= "megAII"
FT FT /function= "AT3"
FT FT 26230..26733
FT FT /*tag= z
FT FT /gene= "megAII"
FT FT /function= "KR3 (inactive)"
FT FT 26998..27258
FT FT /*tag= aa
FT FT /gene= "megAII"
FT FT /function= "ACP3"
FT FT 27393..28590
FT FT /*tag= ab
FT FT /gene= "megAII"
FT FT /function= "KS4"
FT FT 28897..29931
FT FT /*tag= ac
FT FT /gene= "megAII"
FT FT /function= "AT4"
FT FT 29953..30477
FT FT /*tag= ad
FT FT /gene= "megAII"
FT FT /function= "DH4"
FT FT 31396..32244
FT FT /*tag= ae
FT FT /gene= "megAII"
FT FT /function= "ER4"
FT FT 32257..32799
FT FT /*tag= af
FT FT /gene= "megAII"
FT FT /function= "KR4"
FT FT 33052..33312
FT FT /*tag= ag
FT FT /gene= "megAII"
FT FT /function= "ACP4"
FT FT 33666..43271
FT FT /*tag= ah
FT FT /gene= "megAII"
FT FT /product= "megalomicin 6-deoxyerythronolide B
FT FT synthase 3"
FT FT /note= "polyketide synthase; encodes AAB82214"
FT FT 22957..24237
FT FT /*tag= ai
FT FT /gene= "megAII"
FT FT /function= "KS5"
FT FT 24544..25581
FT FT /*tag= aj
FT FT /gene= "megAII"
FT FT /function= "AT5"
FT FT 26230..26733
FT FT /*tag= ak
FT FT /gene= "megAII"
FT FT /function= "KR5"
FT FT 26998..27258
FT FT /*tag= al
FT FT /gene= "megAII"
```


FT misc_feature /function= "ACP5"
FT 27393..28590
FT *tag= am
FT /gene= "megAIII"
FT /function= "KS6"
FT 28897..29931
FT *tag= an
FT /gene= "megAIII"
FT /function= "AT6"
FT misc_feature 29953..30477
FT *tag= ao
FT /gene= "megAIII"
FT /function= "KR6"
FT misc_feature 31396..32244
FT *tag= ap

Query Match 67.7%; Score 17.6; DB 22; Length 47981;
Best Local Similarity 83.3%; Pred. NO. 92;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ccgtctaggcgttcggtccggc 25
|||||
Db 5681 CCGCCAGCCGTCGGTGACCGC 5658

RESULT 5
AAH41415/C
ID AAH41415 standard; cDNA; 1680 BP.

XX AC AAH41415;

XX DT 23-AUG-2001 (first entry)

XX DE Human pentraxin I encoding cDNA SEQ ID NO:1.

XX KW Human; pentraxin I; pentraxin receptor; neuronal disorder; inhibitor;
XX KW detection; neuroprotective; nontropic; cerebroprotective; cytostatic;
XX KW gene therapy; acute head trauma; multiple sclerosis; spinal cord injury;
XX KW Alzheimer's disease; brain tumour; stroke; neuroprotective; ss.

XX OS Homo sapiens.

XX PN EF1101820-A1.

XX PD 23-MAY-2001.

XX PF 15-NOV-1999; 99EP-0122666.

XX PR 15-NOV-1999; 99EP-0122666.

XX PA (MERE) MERCK PATENT GMBH.

XX PI Messeguey Peypoch R, Roselli Vives E, Martinez Escola JM;

XX PI Rodas Gubern B, Adan Plana J, Puig CN, Carceller RA;

XX PI Masa Alvarez M, Piuats X, Den Daas I, Trullas OR;

XX PI DeGregorio-Rocasolano BN;

XX DR WPI; 2001-357928/38.

XX PT New nucleic acid encoding the human Pentraxin receptor useful for
PT preventing, treating, ameliorating or detecting a disease associated
PT with a neuronal disorder such as a stroke, multiple sclerosis, or
PT Alzheimer's disease

XX PS Example 2; Fig 17; 53pp; English.

XX CC The present invention describes a nucleic acid (I) encoding the human
CC pentraxin receptor (hPr) or a protein exhibiting biological properties
CC of hPr and having a sequence selected from: (a) a fully defined 1853 bp
CC sequence given in AAH41415; (b) a sequence which hybridises with (a);
CC (c) a degenerate variant which deviates from (a) or (b); and (d) a
CC fragment, derivative or allelic variation of (a), (b) or (c). (I) has
CC neuroprotective, nontropic, cerebroprotective and cytostatic activity,

CC and can be used in gene therapy. Pentraxin I, hPr, or hPr-encoding
CC sequences are useful in the preparation of pharmaceutical compositions
CC for preventing, treating or ameliorating a disease associated with a
CC neuronal disorder such as stroke, acute head trauma, multiple sclerosis,
CC spinal cord injury, Alzheimer's disease or brain tumour, or for
CC providing a neuroprotective effect. The present sequence represents the
CC human cDNA coding sequence of pentraxin I, which is used in an example
CC from the present invention.

XX SQ Sequence 1680 BP; 328 A; 571 C; 526 G; 255 T; 0 other;

Query Match 66.2%; Score 17.2; DB 22; Length 1680;
Best Local Similarity 86.4%; Pred. NO. 1.3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 cgtctaggcgttcggtccgg 24
|||||
Db 519 CGTCTCGCGCGCGTCTCCGG 498

RESULT 6
ABA46067/C

ID ABA46067 standard; DNA; 381 BP.

XX AC ABA46067;

XX DT 01-FEB-2002 (first entry)

XX DE Human breast cell single exon nucleic acid probe #4762.

XX KW Human; microarray; single exon probe; gene expression; breast;
XX KW disease; cancer; ss.

XX OS Homo sapiens.

XX PN WO200157271-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00662.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-496933/54.

XX PT New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes

XX PS Claim 1; SEQ ID NO 4762; 327pp + sequence listing; English.

XX CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater

PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-488901/53.
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human cervical epithelial cells -
 XX
 PS Claim 25; SEQ ID No 4804; 487pp; English.
 CC The present invention relates to human single exon nucleic acid probes
 CC (SENP). The present sequence is one such probe. The SENPs are derived
 CC from human HeLa cells. The SENPs can be used to produce a single exon
 CC microarray, which can be used for measuring human gene expression in a
 CC sample derived from human cervical epithelial cells. By measuring gene
 CC expression, the probes are therefore useful in grading and/or staging
 CC of diseases of the cervix, notably cervical cancer.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 381 BP; 92 A; 130 C; 95 G; 64 T; 0 other;
 SQ
 Query Match 65.4%; Score 17; DB 22; Length 381;
 Best Local Similarity 80.0%; Pred. No. 1.5e+02;
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 2 ccgtctaggcgttcggtgcggcc 26
 Db 241 CGGGCAGGCGTTCAGAGTCGGCC 217
 RESULT 12
 AAI36227/C
 ID AAI36227 standard; DNA; 381 BP.
 AC AAI36227;
 XX 17-OCT-2001 (first entry)
 DT Probe #4913 used to measure gene expression in human placenta sample.
 DE Probe; microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder; ss.
 KW Homo sapiens.
 OS
 XX WO200157272-A2.
 PN 09-AUG-2001.
 PD 30-JAN-2001; 2001WO-US000663.
 PF 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-488897/53.
 XX

PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -
 XX
 PS Claim 25; SEQ ID No 4913; 654pp; English.
 XX The present invention relates to single exon nucleic acid probes (SENP).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders.
 XX Sequence 381 BP; 92 A; 130 C; 95 G; 64 T; 0 other;
 SQ
 Query Match 65.4%; Score 17; DB 22; Length 381;
 Best Local Similarity 80.0%; Pred. No. 1.5e+02;
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 2 ccgtctaggcgttcggtgcggcc 26
 Db 241 CGGGCAGGCGTTCAGAGTCGGCC 217
 RESULT 13
 AAI04658/C
 ID AAI04658 standard; DNA; 381 BP.
 AC AAI04658;
 XX 09-OCT-2001 (first entry)
 DT Probe #4649 used to measure gene expression in human breast sample.
 DE Probe; human; breast disease; breast cancer; development disorder; ss;
 KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
 KW Homo sapiens.
 OS
 XX WO200157270-A2.
 PN 09-AUG-2001.
 PD 29-JAN-2001; 2001WO-US000661.
 PF 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-476286/51.
 XX Novel single exon nucleic acid probe used to measuring gene expression
 PT in a human breast -
 PT
 PS Claim 25; SEQ ID No 4649; 322pp; English.
 XX The present invention relates to novel single exon nucleic acid probes.
 CC The present sequence is one such probe. The probes are useful for
 CC measuring human gene expression in a human breast sample, where the probe
 CC hybridises at high stringency to a nucleic acid expressed in the human
 CC breast. The probes are useful for predicting, diagnosing, grading,
 CC staging, monitoring and prognosing diseases of the human breast,
 CC particularly those diseases with polygenic aetiology. The diseases
 CC include: breast cancer, disorders of development, inflammatory diseases
 CC of the breast, fibrocystic changes, proliferative breast disease and
 CC non-carcinoma tumours.

CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytostatic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing
 CC inactive proteins or to supplement the patient's own production of P.
 CC Additionally, N may be used to produce the colon cancer-associated ps,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
 CC and AAB77789 represent sequences used in the exemplification of the
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.
 XX
 SQ Sequence 2443 BP; 612 A; 571 C; 599 G; 659 T; 2 other;

Query Match 65.4%; Score 17; DB 22; Length 2443;
 Best Local Similarity 80.0%; Pred. No. 1.6e+02;
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 Qy 2 ccgtctaggcgttcggtgtccggcc 26
 ||| |||| ||| |||| ||||
 Db 130 ccgmctagccgtcggtgccaggcc 154

Search completed: July 31, 2002, 14:11:22
 Job time: 17379 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 14:01:44 ; Search time 6034.22 Seconds
(without alignments)
90.167 Million cell updates/sec

Title: US-09-899-718A-5
Perfect score: 26
Sequence: 1 cccgtcctaggcgttcgtgtccggcc 26

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htgo_inv.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
------------	-------	-------	--------	----	----	-------------

RESULT 1
AX349067
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
Source

AX349067
Sequence 5 from Patent WO0202785.
AX349067
AX349067.1 GI:18615102
synthetic construct.
synthetic construct.
artificial sequence.
1 (sites)
Sprunck, S., Kluth, A., Becker, D., Luetticke, S. and Loerz, H.
Promoters of gene expression in plant caryopses
Patent: WO 0202785-A 5 10-JAN-2002;
Aventis CropScience GmbH (DE)
Location/Qualifiers
1..26
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Oligonucleotide"
1 a 10 c 9 g 6 t

ALIGNMENTS

1	26	100.0	26	6	AX349067	Sequence	5 from Patent WO0202785.	26 bp	DNA	linear	PAT 06-FEB-2002
2	26	100.0	3785	6	AX349063	AX349063					
3	19.8	76.2	30150	1	SC6D10	SC6D10					
4	18.6	71.5	12441	1	AF079762	AF079762					
5	18.6	71.5	37969	3	AC005928	AC005928					
6	18.6	71.5	41944	3	AC005929	AC005929					
7	18.2	70.0	1432	4	CF096127	CF096127					
8	18.2	70.0	1466	4	AF045773	AF045773					
9	18.2	70.0	5908	1	AF147704	AF147704					
10	18.2	70.0	22396	1	SCF43	SCF43					
11	18.2	70.0	173434	2	AC098934	AC098934					
12	18	69.2	3149	1	SLMELOPER	SLMELOPER					
13	18	69.2	6927	1	AF319543	AF319543					
14	18	69.2	20633	3	AC109873	AC109873					
15	18	69.2	40064	3	LMFL6586	LMFL6586					
16	18	69.2	45653	2	AC105239	AC105239					
17	18	69.2	61843	8	NC15E6	NC15E6					
18	18	69.2	79133	8	NC15E11	NC15E11					
19	18	69.2	122393	2	OSJN00101	OSJN00101					
20	17.8	68.5	11355	1	AE008357	AE008357					
21	17.8	68.5	12045	1	AE009255	AE009255					
22	17.6	67.7	41097	1	AF016585	AF016585					
23	17.6	67.7	47981	1	AF263245	AF263245					
24	17.6	67.7	47981	6	AX112026	AX112026					
25	17.6	67.7	165812	2	AP001141	AP001141					
26	17.6	67.7	166063	9	AP002515	AP002515					
27	17.6	67.7	168213	2	AC103175	AC103175					
28	17.6	67.7	168308	2	AC084775	AC084775					
29	17.6	67.7	175464	2	AP002887	AP002887					
30	17.6	67.7	185800	2	AC079366	AC079366					
31	17.4	66.9	115948	2	AC094962	AC094962					
32	17.4	66.9	146722	2	AL139254	AL139254					
33	17.2	66.2	1668	9	HSU48252	HSU48252					
34	17.2	66.2	1680	6	AX137782	AX137782					
35	17.2	66.2	1680	6	AX148704	AX148704					
36	17.2	66.2	2069	9	BC018289	BC018289					
37	17.2	66.2	2070	9	HS0600858	HS0600858					
38	17.2	66.2	2286	9	AF006013	AF006013					
39	17.2	66.2	2396	9	HSU49262	HSU49262					
40	17.2	66.2	3403	9	HSU75651	HSU75651					
41	17.2	66.2	5071	9	HSU61849	HSU61849					
42	17.2	66.2	5146	9	D86963	D86963					
43	17.2	66.2	5966	1	SAMF18862	SAMF18862					
44	17.2	66.2	189050	1	AL646077	AL646077					
45	17.2	66.2	193068	8	AP003455	AP003455					

Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrel@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK

Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics. Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.
(URL: http://www.sanger.ac.uk/projects/S_coelicolor/)
CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).

Stand). The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS. Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon.

where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot program of Alb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.gov/jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

upstream initiation coordin.
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.

Cosmid 6D10 overlaps with cosmid C61 on the AseI-C genomic restriction fragment.

```

FEATURES
  source      1. .30150
  Location/Qualifiers
restriction fragment.

```

```

I. .30130
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"

```

```
/strain="A3(2)"
/ab vref="taxon.100226"
```

```
/db_xref="taxon:1002"
/clone="cosmid_6D10"
```

```
1 823
/clone=
```

```
1. .823
/cond="SC6D10 01"
```

1 923 /gene=";

1. .823
/ccccc=11ccccc10 01"

/note="SC6D10_01" is in conflict with the conceptual translation; SC6D10_01, possible sugar-binding protein (partial), len: > 273 aa; similar to TR:CAB49101 (EMBL:AJ248283) *Pyrococcus abyssi* maltose-binding periplasmic protein precursor MalE, 453 aa; fasta scores: opt: 246 z-score: 288.7 E(): 1.1e-08; 24.0% identity in 271 aa overlap and to SW:MAXL SPFRN (EMBL:L08611) *Streptococcus pneumoniae* maltose/maltodextrin-binding protein precursor MalX, 423 aa; fasta scores: opt: 184 z-score: 217.6 E(): 0.0001; 22.9% identity in 279 aa overlap"

```
overlap
/codon_start=2
```

```
/column_start=2
/transl_table=11
```

```

/transf_cable-11
/product="putative sugar-binding protein"

```

```

/product="pulcrative sugar
/protein id="CAB71196.1"

```

```
/protein__id=CAB71190
/db xref="GI:6855382"
```

Y0D_Xref="GL:0003502
/translation="MKAPPTKSEFAQVAKKLTUKTGQYRQLQGLMPNHYIGYTVVDY
YMSQWDHAYFDGCRSAVEDAEYFGYKQLVDDLGGEKLEKGYRTNFTGDGWAQ
HFQQTQYQVLQDGEWPLFMKADAGVDFEIGTAVPVADEAAEYKRGYSFTVMGIA
PQKQONAAWELVKTITDGTAVAVAFANAIRNVPFTFALKSPDLKTDPEKFTLIDIA
HQRESNPSPASVNGATYQLTLDLQIGYQYEGSKVKDLKAGLEKAAQYLDRIEQAQ"
1 100
signature

feature

```
/gene="SC6B10.01"
/note="nominal ov"
```

```

/note="nominal overlap with S. coefficient cosine
830 1774

```

830. .I774
/gene="SC61

```
/gene="SC6B10.02"
```

misc feature

gene

Query Match	100.0%;	Score 26;	DB 6;	Length 26;
Best Local Similarity	100.0%;	Pred. No. 0.53;		
Matches	26;	Conservative	0;	Mismatches 0;
			Indels	0;
			Gaps	0;

Qy 1 cccgtctaggcgttcggtgtccggcc 26
|||||
db 1 CCCGTCtaggcgTTCGGTGTCCGGCC 26

db 1 CCCGCTAGGCGTTCGGTGTCGGCC 26

RESULT	2	DNA	linear	PAT 06-FEB-2002
LOCUS	AX349063	3785 bp		
DEFINITION	Sequence 1 from Patent WO0202785.			
ACCESSION	AX349063			
VERSION	AX349063.1	GT:18615098		

```

source
1. 3783
/organism="Triticum aestivum"
/db_xref="taxon:4565"
945 a 980 c 899 g 961 t
BASE COUNT

```

Query Match	100.0%;	Score 26;	DB 6;	Length 3785;
Best Local Similarity	100.0%;	Pred. No. 0.44;		
Matches	0;	Mismatches	0;	Indels
26;	Conservative			

QY 1 ccgcgtctagcgcttcggtgtccgccc 26
|||||
pb 2107 CCCGCTAGCGCTTCGTTGTCGGCC 213

db 2107 C C C G T C T A G G C G T T C G G T G T C C G G C 2132

RESULT	3
LOCUS	SC6D10
DEFINITION	Streptomyces coelicolor cosmid 6D10.
ACCESSION	AL138538
VERSION	AL138538.1 GI:6855381
KEYWORDS	ABC transporter ATP-binding protein; globin; kinase; lipoprotein;
	linear BCT 01-FEB-2000

SOURCE	ORGANISM
Flora, feces	<i>Streptomyces coelicolor</i> A3(2).
Flora, feces	<i>Streptomyces coelicolor</i> A3(2).
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.	

TITLE A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb *Streptomyces coelicolor* A3(2) chromosome

JOURNAL Mol. Microbiol. 21 (1), 77-96 (1996)


```

/product="putative tautomerase"
/protein_id="AAC68676.1"
/db_xref="GI:3789893"
/translaton="WTDDLGTALTOPPLGRVTRAVADRELCTHLETRGHIWHAAG
DPAVYVLRGQADDPYPAERYVRARGALSFSTGSGVAVVADHAAALSLCSDFGSGAD
GVFVQVLSGEGCPKEREQVLPFAAGDPEGQGVVVEGTHRETLGLADPDSASTA
FELLGTRVPVAVTAAAVLGVVADRRADFADLLERPLSDSLAPOSLSRTVRAAG
ALLETALLADSDSPGALSALGVTAQVLTGNVALLALAHPEQWRELCDPGLAAA
AVETLRYPVQIDARVVRGTELAGRLRFPAGAVVVLTAATGRDPEVFTDPRFDL
APDAAHALHAPGAPYGPVASLVRLQAEVALRTLAGREFGLRQAGDVLRPRAPVG
GPLSPVPSSS"
1475..2755
/gene="desVII"
1475..2755
/gene="desVII"
/codon_start=1
/transl_table=11
/product="glycosyl transferase"
/protein_id="AAC68677.1"
/db_xref="GI:3789894"
/translaton="MRVLLTGFHHTHYGLVPLAWALLAAGHEVRVRSQPALTDIT
GSLAAPVGTGDLHLEYVRMGPRNHPAIAFDEARPEPLDWDHALGTEAILAPY
FLLANNDSVDDLVDFARSQPDVLWEPITYAGAVAAQVYGAHARVLWGPDVMS
ARKKVALRDVPPHREDPTAEWLTWLDRYGASFEEELTGQFTIDPTPPSLRDT
GLPTVMRYVNGTSPVVDMLSEPPARPVCLTLGVSAREVGLGDGVSQGDILEALA
DDIELVATLDASQAEIRNPKHRTFTDFVPMHALLPSCSAIITHGGAGTYATAVIN
AVPQMLAELMDAPYKARAVAQGAGFELPPAELTPQAVRDVAVRILDDPSVATAAHR
LREETFGDTPAGIVPELERLAAQHRPPADARH"
2845..3558
/gene="desVI"
2845..3558
/gene="desVI"
/codon_start=1
/transl_table=11
/product="N,N-dimethyltransferase"
/protein_id="AAC68678.1"
/db_xref="GI:3789895"
/transl_table=11
/product="N,N-dimethyltransferase"
/protein_id="AAC68679.1"
/db_xref="GI:3789896"
/translaton="MYEVDHADVDYDLFYLGRGKDYAAEASDIADLRVSRTPPEASLLD
VACGTGLEHTKEFTAGLESEDLMTAKRKLDPATLHQGDMRDFLGRKFSAV
VMFSSVGLYKDTTELGAASFAEHLPGGVVVEPWPFETFDAGWVSDVVRDG
RIVARSHSVRGNATREMHFTVADPGKGVHRHSDVHLITLFLHQAEEAAFTAAGLR
VEYLEGPGSGRLGVGPA"
3625..6054
/gene="desR"
3625..6054
/gene="desR"
/notes="putative extracellular protein involved in a
resistance mechanism"
/codon_start=1
/transl_table=11
/product="beta-glucosidase"
/protein_id="AAC68679.1"
/db_xref="GI:3789896"
/translaton="WTGKTRIPRVRRGRTTPRAFTLAVVGTLLAGTTTVAAPGAADT
ANQVTSRAELVAQMLTDEKISFVHWALDDPDNRVNGVPGVPLGIPELPAACDPCNG
IRLGGTATLPAVALASTFDDTWMADSTGKVMGRDGRALQMDVGLGPMNNIRVPHG
GRNYEFSDDPLVSRSTAQAQKGGAGLMTAKHFAANNOENNRFSVNAVNDQTL
RIEFPFAFSAKGAAGAACFMYAGNLGKPCSGNDELNNVLRTQWFGQVWMSDLA
TPGDAITKGLQDMGVELPGDVPKGPPAKFFGALKATVNLGTVPEAAVTSRGS
RIVGMEKEGGLLATPAPRPERDKAGAAVSRKVAENGAVLLRNEGQALPLAGDAGKS
IAVIGPTAVDPKVTGGLSAHVVPDSSAAPLDTIKARAGATVYETGEFTGQIIPA
GNLSPFNOCHLEPGCKAGALXDGLTVPADGEYRIAVRGGYATVOLGSHTEIAGQ
VYGVSSDPLKTLTKTHLITSGFAMSATPLSLGELGWVTPPAADATIAKAVESAKAR
TAVFAYDGTGVRPNLSLEGTQDKLISAVADANPNTIYVLTNGSSVLMPLWSKTR
AVLMDYTPQAGAEATAALLYGDVPSGLTQSPFAAENQHAVAGDPTSYFVDNQOT
YREGIHVGRWFDKENKPLFPFGHLSVTSFTSQAPTVPRTSTGGLKVTVTVRNSGK
RACQEVQAYLGLASPNVTAPOAKKLVGYTKVSLAAGEAKTVTVNVDRQLQTGSSSA
DLRGATVNVW"
complement(6143..7282)
/gene="desV"
complement(6143..7282)
/gene="desV"
/codon_start=1
/transl_table=11
/product="4-dehydratase"
/protein_id="AAC68684.1"
/db_xref="GI:3789901"
/translaton="MKSALSDLAFFGSPAAFDQPLLVGPNRIDRLARLYERLDRLDLS
QWLSNGGPLVREFEERVAGLAVATCNATAGLQLLAHAAGLGEVIMPSMTFAA
TPHALRWIGLTPVFADIDPDTGNLDPDQVAAAVTPRTSAVVGVHVGRCPCADQLRKY

```

ADHGLRLFDAAHALGCAVDGRPAGSLGDAEVFSPHATKAVNATEGGAVVTDADLA
ARLALHNFGLDLPGGSPAGGNTAKMSEAAANGLTSLDAFPEVIDRNRNHAAYREH
LADPLGVLYADHRHNNHQYVIVEIDATGCIHRDLVMEVLKAEVHTRAAYFSPGC
HELEPVGQPHAPLPHTERLAARVLSLPTGTAIGDDDIRRVADLLRLCATRGRELTAR
HRDTAPALAAPOTSPTTIGRSR"
BASE COUNT 1704 a 4294 c 4686 g 1757 t
ORIGIN

Query Match 71.5%; Score 18.6; DB 1; Length 12441;
Best Local Similarity 84.0%; Pred. No. 7e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 cccgtctaggcgttcgtgtccgcg 25

Db 11129 CGCGGAGCGGTCGCGTGTGCGGC 11153

RESULT 5

AC005928
LOCUS Leishmania major chromosome 3 clone L6290 strain Friedlin, complete
DEFINITION
ACCESSION AC005928
VERSION AC005928.4 GI:7025823
KEYWORDS
SOURCE Htg.
ORGANISM Leishmania major.
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
1 (bases 1 to 37969)
Myler,P.J., Sisk,E., Hixson,G., Kiser,P., Rickel,E., Hassebrock,M.,
Cawthra,J., Marsolini,F., Sunkin,S. and Stuart,K.D.
Direct Submission
Submitted (04-NOV-1998) Seattle Biomedical Research Institution, 4
Nickerson Street, Seattle, WA 98109-1651, USA
2 (bases 1 to 37969)
Myler,P.J., Sisk,E., Hixson,G., Kiser,P., Rickel,E., Hassebrock,M.,
Cawthra,J., Marsolini,F., Sunkin,S. and Stuart,K.D.
Direct Submission
Submitted (15-NOV-1999) Seattle Biomedical Research Institution, 4
Nickerson Street, Seattle, WA 98109-1651, USA
3 (bases 1 to 37969)
Myler,P.J., Sisk,E., Hixson,G., Kiser,P., Rickel,E., Hassebrock,M.,
Cawthra,J., Marsolini,F., Sunkin,S. and Stuart,K.D.
Direct Submission
Submitted (24-FEB-2000) Seattle Biomedical Research Institution, 4
Nickerson Street, Seattle, WA 98109-1651, USA
COMMENT On Feb 24, 2000 this sequence version replaced gi:6425646.
FEATURES
source
1. .37969
/organism="Leishmania major"
/strain="Friedlin"
/db_xref="taxon:5664"
/chromosome="3"
/clone="L6290"
660. .3290
/gene="L6910.5"
/notes="L6290.4; predicted using Glimmer, Testcode and
CodonUsage; ORF starts at 609, Glimmer predicts start
codon at 618"
660. .3290
/gene="L6910.5"
/codon_start=1
/product="L6910.5"
/protein_id="AAF35925.1"
/db_xref="GI:7025826"

gene

/note="predicted using Glimmer, Testcode and
CodonUsage; ORF starts at 609, Glimmer predicts start
codon at 618"

CDS

660. .3290
/gene="L6910.5"
/codon_start=1
/product="L6910.5"
/protein_id="AAF35925.1"
/db_xref="GI:7025826"
/translation="MTNTCCGFAAPLPHPAVSSPSCQCPSTPTIDTLARQLHQDPSP
QPKQYWRERRSSSLAVSPSSASSASHSTDAYEVAVVAETRRDVGSLIQOP
TSPCTVKSYGTASD1SDGSRSSLSAVVHTVANTFDLMTOTVGYRTSTAEV
QFQLEHQAHQFQORQGVYSAFEFSDHSHFTLLDGNAMVGGVGGSPAPLAR
DTSQEPLEGGANVAGPAEGRFPASGNGVGPAPVFLPRSAQSPGPPPTF
VELVMRVGSTSLQFLMWVLLVLLVGNAFQVIFLNFHQTFLNPNQIAPASS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

SASSGSGTSSSGRHAELASSYTFVISAVLFPAFFVLLITYALWRRRNLFRTREW
AGWLVFVIGAMDALNSALFYAAANTPEVLQALFVSLVPIYSATFKWLLKOPRDYA
NPYVVFVLLATVALASFNQVTHHHHTEDSRSGSHGACVKELLVAVSDGSSSL
SPVALDRFLWCFIFELSPVPTVLANVQTMVIRYTNDDOLMAYLAHADAEAGCESS
NDVAAGPQSTLLDSDARDSHGDAHCPSPAAGEAVERVPLQPSVVSRSPLPHGW
HQQHHPHSHLLHGWDASVKLMLTSDTAQIALVAVIMPMDALPWFVGSHSRTVQV
NLDEGIDCVLHCPNRMRYCILYSTGFVLYTASAYLNRYSVTLCSMYSQLSGPTIALY
LIAFPTLMTDDASPYYSVEAIFLLSGTVLYYWDVMTVEEKAVGEMOLKWMQOE
QSPRHAPSLAEGORYHEVDGNGSSSLATSGSGVSPQQQQOHRHARSRRYCRRRQSGY
AVVVDQASDASASQPHRY"
4472. .6754
/gene="L6910.6"
/note="predicted using Glimmer, Testcode and CodonUsage;
Glimmer predicts start codon at 4496"
4472. .6754
/gene="L6910.6"
/codon_start=1
/product="L6910.6"
/protein_id="AAF35926.1"
/db_xref="GI:7025827"
/translation="MSPIHRKHMEVLGLGVGTANDASNTORPAGALSREEOQLPRES
SLTTCATAAAAAAASGDVRHCPLEDDLSNATVLAATAETATRPPLFGSLRPF
SDALASPLRHLIRGSSFPASPLRITSVCLEPCHTGAETAAATRTAFAHAGANV
VLMRSLPRTASVVMPSLRAFRERLCKMSLTQILWILLVALVMGNFQIIVMLNF
LISFSDGTPGNYTAFAYGFFAVFLVILGAYTAIRRPSLRFARHAHGVILIGV
FCDAFNSWLTAAASYTSEVLQALFNLCPYAVFLCKWILHDTRRYANVYIVSVFL
TICGILAAISLYGLIKDHNMGEKWWILIFLSPMPFVLMVNMWQSLWIVYTHDPNFVL
WLMRELATAEAAEPVSAARGSEVTPDTSDFRGADVAGPASPADADCAKVAS
SIARYNDRTLSGLEDDEDDKEELRNARPARHAKTKLANSVKPSSVDANDDAVGDS
EAHEPVASDSDSGTAGSVAITPAASDEPATSIRVNEHGVHDLSGVEDGEYLRPTAT
GAPLPTMFTVRYNOGEDTVVVKMLAGETVQLCTFLLCLPADALPMMKSETVS
ATLNDFAEGRCVFTIRDNFLYCFLYTIGFVTVGCAYLNHYSAALCSITVQLSSV
TALMLVIPSNMVQVGDSPWYNMFAIFLCLVAALIVVMWEENTDAEKVQAEVELKM
KELVRSSSHEAPHLYTMQ"
7317. .9026
/gene="L6910.7"
/note="L6290.6; predicted using Glimmer, Testcode and
CodonUsage; Glimmer predicts start codon at 7404"
7317. .9026
/gene="L6910.7"
/codon_start=1
/product="L6910.7"
/protein_id="AAF35927.1"
/db_xref="GI:7025828"
/translation="MRGCKHKLARSFOQWKLYNALCDROMPTPLHATIRLPVPR
VAHPPSCPSAPASYDMLQNTNSSAVLQWQREYDFLLGCDASAHRTAGTPPATTA
THVIASTRVSAHVGPAPPLIGYPDYIAIEATQHTEGVIRGETYAHAAAVGAVPAPP
PGAAPSPVKWIIQPLLDGFVAQRLTAHIGVTTGSLDTRRTANALGRQLPPVELSPF
YAASALLDRWCLFGEAAAVVPSMPASGATLDGQQQRPHDAHHLRIAEALGAF
LLSNVRAVLNGAVLSHPGTARSVLYVPPRRSGKTLALHCLSAAMAQSSASGDDG
VRLTAAEHFFLAACMPVRRMLDPASTRLTSTPTIFACSLPHHITVGLGAVLGRPNP
LAAAVELPSFLQCEAGLACLRLNDTGVLDMSKTYRVRLIEDIYRASEDSISTQWAPV
VNTLAGVLLDWNVEELASPTVRNAREIPLHDGVEELKRAQDYMFPHGHLLR
TVYDAPQDAPARLADALADEWDTPAQRSLANGTAPALHCTEGSVNFDLATLMSLLS
ERA"
19200. .20060
/gene="L6290.1"
/note="predicted using Glimmer, Testcode and CodonUsage;
Glimmer predicts start codon at 19287"
19200. .20060
/gene="L6290.1"
/codon_start=1
/product="L6290.1"
/protein_id="AAF35928.1"
/db_xref="GI:7025829"
/translation="MSSHQILITAPEYKRFQAOLEARSLIPMAVPIETIATPDPTA
DMRALLODLNGIDYIAFCSCAVDSLSAALDHYRGGGSGVSEAGSAPVNEEDRA
TAVYVLAKALGAIGADYVFERLRITTCIDPEPSAGIAAKLAESRAGERTTA
VLAPCVGTEPDVVPREVQOLKGMKVVVDAYMTRPVSEEAASVAVTALNGSVR
GVATISGEIVALLRRSPQCLMNDIVACFGPYTAGFAAKHGVKVCVAKOFRSFDGFA
ATIEQHVASR"
complement(22336. .28971)
/gene="HEL2"

/note="L6290.2; L6910.1L; predicted using Glimmer,

```
CDS
Testcode and CodonUsage"
complement(22336..28971)
/ene="HEL2"
/note="blastp similarity to helicases"
/codon_start=1
/product="Helicase 2"
/protein_id="AAF35923.1"
/db_xref="GI:7025824"
/translation="MSSHYRELNLQFVNDANGQPDFATVETLSLPIKOLFQERSLLC
PSSSEATASNASPTVAPVHETITALLTCNVLIRWAAERAGQDQGAARAA
AVQSDLSQFDEGSRMAKFLATFQAPALVCIDKALTEVADFHYAARQFDMF
VHLSMERKFFIDGSLMALSSPHVDMLQPLHLHNAQVLLRGLQSRGHEHIF
FRHTAWFEPPOKLEFREALRFLTAATNPDSRLVDFDFSFSEFAFVSKYE
PEFLMSDEOLGHELMFLHNTHEELLREKEGEGEGTGPASASAPLTDDE
KLELQRTEDQTMELKRPYRSLVDESDEGDTVAFYHCFLTWATARRIKVAYSRII
TRENVAVFVYSSGFARSILFENPSTKLAALERAVPLTUSDAALAKDKGOLS
REQVVSAAHAYLHAGDSPPTEQOQLQCVMTTAYCTALVSALRAQSVTSVP
VAAFNLQAFVQDCATGTEFDVYDGHILITIMHLLRTTPAEQLLDEGI
TDVNLSDQVVDGSAIITGGALSALPDVLAEPKAMKYPHLTHDILQILARFVE
GHVSNRPYPATLGNALGCVIYVQPFDELNDVDVAVVERESQNTYTERKQLAME
AAFTDAAQAQESMGIRGFTAGTLAVVCSDESDEDAADTGAAGTSAGKVAKYKQ
HSGKNKAKAREDIIREENLRDANKTVQDWTQVQAECLKTKDQASRLTVADSLG
LFAATIRMTGTSFGNDDPGEVLKDSAPVQLQTMWRLVLDASGMREAEIATITDD
ADRSAPADKRRKSAKDRATNLYAFGLIADLEENITNAKSGSHGSDMGLELRK
VKGDTFVYDASTYLNWLYLYVOLHPQCKLMTVVRLQLMQWNERERAKQOAPNI
NVGVPLEYCHHKVLACIRNGVAIISTDDLDAVSALAHDFPASYEKVDAIGAKQ
NSAVSVLPSPHPRKACFTPEMOLMTMGHLERPVIPTKDYRVFNPDCWORELL
DIVNRGSAVCAPISAGKFTFISYCWISALTKSNKVVYVAPNRLINQAVDYCA
RYGSKSFPFNVGHGADHYRVDSQVLTLPVLETLILLSPKYEKAKRLDY
LIDTHITMSSGNGYMERVALLPCPEVALSTGTEQQLCGWLNVRQLEAQOP
EAAKRDFTVHDIPSSGSIQWINDIKKMYLPPPGYQPKLTKAKYPIHDLPLS
ILTLQLOSQPPDITLVPEVSVOLYOKMCTLTFEHVCDAGKVAIVQAMKAQLMSLL
PETYEQIYITQORAEQADYNAFVYVNLKGCDCDGLFELSGEVVAENTDMR
ELCETILQSQRLEDEAQLDKYDAALKPEGRPDTSAEAEATASATATATTA
IAQPKAFPEKSEFIADNIIISRELNRDMGPTVTFEGEDCDLTKAILARDED
AAAYATQEEPAKATMERKAAAEARRQESTLKQKLTLDGDEGVEVADVDN
NNNDEVDVDPVLEPTEFTIGQYTDVTVVKEAVEECIKRGDTLCARAIQRGVHH
AGVKGKLRMEILFRGHGVCATETLALGHSPCRSVVLADGHLVLLNTVQFQRM
GRAGRGDLGLHILFSLVSMKKITRLTSMVTIKGNVQVDPITOLRLQLYDVARH
RLDKAESWRGALGAERLYVNPFFAGRAMENGMSFTVEFAQLMLGLFHRGL
HFKDRPSSGLLDVDMVFRNAQGVGFATISMITRKYFAKARLRFRYLAIPSD
AVVQAKALLAYLFSVKNYCVPELHRSALSDPAVIELWSRRVAKRFRYLAIPSD
VCSPATKPSVSVFMSAFYTYLASTLKEPADTRLPYMCATKKKHCIFGVTAMD
LMAKLEAPYAADPFAIDVSCGDFHVSDELVTTLRGLFCDRMLPLVDLTDGC
RIDGAQLINACITDFVRCQADISVRSNFRTELLELNGLSQSEYAVLNQTERLS
NISGLTRDKYTPLOVLSALYPERGENVAEDLKFTAGTLGVMDFASQLHELPHI
DKELAVQWQARIEMRRRK"
complement(29757..>37969)
/ene="L505.2"
/note="L6290.3R; predicted using Glimmer, Testcode and
CodonUsage; Glimmer predicts start codon at 15410"
complement(29757..>37969)
/ene="L505.2"
/codon_start=3
/product="L505.2"
/protein_id="AAF35924.1"
/db_xref="GI:7025825"
/translation="SRHRERHWEPSASATTLAADAAPQSAATSSPHVHEDTTA
TSSPAPPTQPSSTISGAGKVKRATKRVGGVAPRAVASSRDSNGNARTSDRSTQ
QQQQQQDDANDSVPTATTPQYNDFOCRULVALGEARRPLTSEGHMALPENADAG
VAVQNPVPSQPLAPPTSAFAPPTATGGSRRARMPLVFWAPRLIASFTIHG
ALCSLELMSQWELTPENWILLIALLFETLNPAGDGRVSDVDAERHELRVTAAG
VDGGSALAAAYSEERHOLGMEYIRRAHPHLFRYVGLPSESARQDGMPSLPSVS
ESRGGLSASNPNTIAAAATGCMVQNRHDMYFDSGAVTATGSDVSASPASHAQ
TATTPSLMGFTGTHGSGSALPHVLRDWSHLNTADALVATAPSEPSVSP
PLCLPTLGQOQPRTPATKAPVWMLDPSALLADAEGRVLFQSGVRGVOQAVSTTAA
GEDKYSAQDSTPTPDATKRVGASPKYSGSAFFPSLPAPLQOQRTPTATPPSSA
STVSFVYMAERVLARPALWHVAYATPHPTAASCVCQVQLQARVABLPTPTLL
PPPSISGTSSTGIPATSAEESQVTOVRLAEGVSCERGAADSAAPLPLEVHL
IGDAGNVQEARVACHLRNVVSATARKFSSTVNPILLNASGSGSPSTQRTSLPSL
QSTAPSPAGDVHCHLPRGLTLHCRNMVLYGHLRVFNLVLDSCFVFGSLTEELAS
VQVSRSHLALMDRLILLRPPYRRGFSRDRGAGRVGRASSTASPGAADARHP
```

Query Match 71.5%; Score 18.6; DB 3; Length 37969;
Best Local Similarity 84.08; Pred. No. 6.7e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ccgcctaggcggttcggttcggc 25
Db 14231 CCGCTGCGGGGTTCCGGCTCGGC 14255

RESULT 6
AC005929
LOCUS
DEFINITION
AC005929
AC005929.5 GI:7025830
VERSION
AC005929
HTG.
KEYWORDS
SOURCE
ORGANISM
Leishmania major.
Leishmania major.
Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
REFERENCE
1 (bases 1 to 41944)
Myler,P.J., Sisk,E., Hixson,G., Kiser,P., Rickel,E., Hassebrock,M.,
Cawthra,J., Marsolini,F., Sunkin,S. and Stuart,K.D.
AUTHORS
Direct Submission
TITLE
JOURNAL
Submitted (04-NOV-1998) Seattle Biomedical Research Institution, 4
Nickerson Street, Seattle, WA 98109-1651, USA
REFERENCE
2 (bases 1 to 41944)
Myler,P.J., Sisk,E., Hixson,G., Kiser,P., Rickel,E., Hassebrock,M.,
Cawthra,J., Marsolini,F., Sunkin,S. and Stuart,K.D.
AUTHORS
Direct Submission
TITLE
JOURNAL
Submitted (15-NOV-1999) Seattle Biomedical Research Institution, 4
Nickerson Street, Seattle, WA 98109-1651, USA
REFERENCE
3 (bases 1 to 41944)
Myler,P.J., Sisk,E., Hixson,G., Kiser,P., Rickel,E., Hassebrock,M.,
Cawthra,J., Marsolini,F., Sunkin,S. and Stuart,K.D.
AUTHORS
Direct Submission
TITLE
JOURNAL
Submitted (24-FEB-2000) Seattle Biomedical Research Institution, 4
Nickerson Street, Seattle, WA 98109-1651, USA
COMMENT
On Feb 24, 2000 this sequence version replaced gi:6425645.
FEATURES
Location/Qualifiers
1. 41944
/organism="Leishmania major"
/strain="Friedlin"
/db_xref="taxon:5664"
/chromosome="3"
/clone="L6910"
1420..2418
/gene="QOR"
/codon_start=1
/product="Quinone Oxidoreductase"
/protein_id="AAF35929.1"
/db_xref="GI:7025831"
/translation="MEVIATAHGGEVLAIVRPSHTPDATQLEGQVILVHNAAGVNF
IDYFUSGLYKKPAMPYIVGEEGAGAVKVGAVPETMLGKRVAYFGGCTGYAAF
TVAPALAREVDPGVDAEAAVNCQGLTAHYLVDSYPCPGSTVVVHAAAGTGILL
VCOMALKRARGIVGCGGAETLARSGRADVVIDYVATPDWAPLVRAAPQGVDAV
VDGGAATAGSLSVLRPRGWTITFGNAGVAPVPSPLELSRAGSVYLQRPITGDMWR
TPEVQRRPSEVFEVWASKQVQLTIGHEYPLHEAAQALADIQSRKTTCKLLKCID"
2880..3416
/gene="L6910.1"
/note="predicted using Glimmer, Testcode and CodonUsage"
2880..3416
/gene="L6910.1"
/codon_start=1
/product="L6910.1"
/protein_id="AAF35930.1"

gene
/db_xref="GI:7025832"
/translation="MKLLKFNVDAPHTLADVQLRNQRLWPCRRHGNAVAAKSWTIT
AASPVSQGYWLLDDSPSTIDVLRQLEEMVQVRRERRVRAQRQIRQAKLMRLQ
DFRSKMGQMERATYATGATSKATKARTMNAEEMTEKKNMKREAREAAEAIED
NDASVMPKRRKERSKK"
5389..9468
/gene="HELI"
/note="predicted using Glimmer, Testcode and CodonUsage;
CodonUsage: ORF starts at 5110, Glimmer predicts start
codon at 5389"
CDS
5389..9468
/gene="HELI"
/note="Blastp similarity to CHL1 helicase"
/codon_start=1
/product="Helicase 1"
/protein_id="AAF35931.1"
/db_xref="GI:7025833"
/translation="MLNLSILSYLFDVEDAKVDGDTASGSGSVVAEPAAATATST
HDEKAALDIPSPSTIDVLRQLEEMVQVRRERRVRAQRQIRQAKLMRLQ
QOLNNGGHEDFLLTODPLAWYBOLSMKKARLGGDDLDVLHSASSLSSSSSSSS
SAREDDDDAQAQLEAALSTLPLPKPKVYFASRTHQLOQLMEDLQRTAFARLP
HRIEPLAESAGRTAGSGMGVNDTDDPATSSKSLGSPTHACPVAAASRTPSRQOQ
PRRLTAVHVRPHLCNLAARLKKAGGNDRNLNLYCREAMHFFERSKQGRYRQOQEH
POQPHPTDRPVRNATEDATGQOELEDDKGCYCVESHLRSLMTYLRDEQRADA
TDPVDAASGASGNGGPPSVYMDRLRLGAELQACPYLATRLLDNGADYAFIPY
GYLDEQORAVLLCGAATPATAAEAATFATSRSTPAAEGGALONDVAPHSPL
GTVLYHRQVATATFRRRQOQORNGVYGGDDDEGEAEIWSMTMARSPSILC
GDLVFDEANIADCRSTVTVPVPHILLARLLETYLARYASRLTRNKQRLREL
VHFLKSLGSCERADSAVLGEGGEGDIAPLSAPSPMLRLPQSTPTPTTAART
LVLPFHFFLDAGIDSDVYAFITFLVDSQLMKLGFVSYALDAELQOQERESYSTAM
RTMTVSGTDSAGTAGVRKQSGVGVSDAHNEAQORRLASLDLPDNHKTAT
EGEAAPHLLSLAEHLSQOLQAGVAMTAAAHAGAVAPADVOLRALTAELQVLERL
CALYVNSTTTRVLWTPSSLSPTSPARPACAVROGALKVIOLEPGMYTEAPLVEAR
AVLAGTQMPLAFTCGPILPAQAMVGDDAGGCGTASTIGRSVAVAVEGIGSDYV
HCIDASQGVVAPGAPGSAFISHEGVHVPSSVQVWALGPGSLKMELSQQALGUR
SNAASTNDGSSRGIVSSISPHARVLAEGVCTLLNARVLPPAGAICFTSYDYM
DAYVLESTGYTAQISDVKRIKFKETRNGGGGGGGGRTAGTGGGGGAEIAEL
LREYQWISGEPASGAAERAPGSTTASSPVSTAASQOROKPSRRGAFPLVAMGG
RLSEINFADLGRVAVLGNVPYANPTDVERMNLKHIVATRLANADARBMGRTA
GSASTPSASSPFTCAEGLWYMDSMRVNQCIGRCIRHAGDYATILLDARTER
PDARRRVSNLQSMRVQAQTGQCFSGVREFFAERQPKG"
10078..11295
/gene="PAMT"
/note="L6910.10; predicted using Glimmer, Testcode and
CodonUsage"
10078..11295
/gene="PAMT"
/codon_start=1
/product="Protein Arginine Methyltransferase"
/protein_id="AAF35932.1"
/db_xref="GI:7025834"
/translation="MPANRMRFGAPTORGESRSANASKVLTESLATDAENKPTSK
SLYGDAAARLSNNIMTHDQRLURAYEYIFRNVRKTLHLGCGMGILYTLAARGMAK
MVTGVSSALVDAARVAEQNGKLNILIRCLVDVHLQPGMKFYDLCWGMPLLL
LNFRVLTALYAAHRLTPSGALCPNRAHLVAVSDYSFRLEDTEFWSNVYFGOMP
MKELVREYEMCAIPGNSIVSAPCLATHMTDLEGLTVEETATYEAQAOAASRON
EENPVHRVPTAVAGQGYEAATLSITRNATVHYLTFFVLDATFSTKTPGNFVLA
VRGGENNTEVSVGLREPLVPVNGEIQGLTIRIYTPADKGGKITVVEVTAKGVQA
AVETSGYIYQSY"
12821..13870
/gene="L6910.11"
/note="predicted using Glimmer, Testcode and CodonUsage"
12821..13870
/gene="L6910.11"
/codon_start=1
/product="L6910.11"
/protein_id="AAF35933.1"
/db_xref="GI:7025835"
/translation="MMSRVRPHIQRISVEDIMKGNRRLAKAITLAESNPADSTR
LGNLRIRHQNATVIRVFPALSGPGSGKSSLDLTUGYLCEKKNHVGVLAVDPS
TITHGSLGTQCKRKLHNSYIRPSPGHLGGVTSRAWEALETFEAREDDVVF
ETVGVGQSTQCKDLTDMILLVPPASGDELQGIKGVWEVADVVMVTKNGRRKPLV

gene
/db_xref="GI:7025832"
/translation="MKLLKFNVDAPHTLADVQLRNQRLWPCRRHGNAVAAKSWTIT
AASPVSQGYWLLDDSPSTIDVLRQLEEMVQVRRERRVRAQRQIRQAKLMRLQ
DFRSKMGQMERATYATGATSKATKARTMNAEEMTEKKNMKREAREAAEAIED
NDASVMPKRRKERSKK"
5389..9468
/gene="HELI"
/note="predicted using Glimmer, Testcode and CodonUsage;
CodonUsage: ORF starts at 5110, Glimmer predicts start
codon at 5389"
CDS
5389..9468
/gene="HELI"
/note="Blastp similarity to CHL1 helicase"
/codon_start=1
/product="Helicase 1"
/protein_id="AAF35931.1"
/db_xref="GI:7025833"
/translation="MLNLSILSYLFDVEDAKVDGDTASGSGSVVAEPAAATATST
HDEKAALDIPSPSTIDVLRQLEEMVQVRRERRVRAQRQIRQAKLMRLQ
QOLNNGGHEDFLLTODPLAWYBOLSMKKARLGGDDLDVLHSASSLSSSSSSSS
SAREDDDDAQAQLEAALSTLPLPKPKVYFASRTHQLOQLMEDLQRTAFARLP
HRIEPLAESAGRTAGSGMGVNDTDDPATSSKSLGSPTHACPVAAASRTPSRQOQ
PRRLTAVHVRPHLCNLAARLKKAGGNDRNLNLYCREAMHFFERSKQGRYRQOQEH
POQPHPTDRPVRNATEDATGQOELEDDKGCYCVESHLRSLMTYLRDEQRADA
TDPVDAASGASGNGGPPSVYMDRLRLGAELQACPYLATRLLDNGADYAFIPY
GYLDEQORAVLLCGAATPATAAEAATFATSRSTPAAEGGALONDVAPHSPL
GTVLYHRQVATATFRRRQOQORNGVYGGDDDEGEAEIWSMTMARSPSILC
GDLVFDEANIADCRSTVTVPVPHILLARLLETYLARYASRLTRNKQRLREL
VHFLKSLGSCERADSAVLGEGGEGDIAPLSAPSPMLRLPQSTPTPTTAART
LVLPFHFFLDAGIDSDVYAFITFLVDSQLMKLGFVSYALDAELQOQERESYSTAM
RTMTVSGTDSAGTAGVRKQSGVGVSDAHNEAQORRLASLDLPDNHKTAT
EGEAAPHLLSLAEHLSQOLQAGVAMTAAAHAGAVAPADVOLRALTAELQVLERL
CALYVNSTTTRVLWTPSSLSPTSPARPACAVROGALKVIOLEPGMYTEAPLVEAR
AVLAGTQMPLAFTCGPILPAQAMVGDDAGGCGTASTIGRSVAVAVEGIGSDYV
HCIDASQGVVAPGAPGSAFISHEGVHVPSSVQVWALGPGSLKMELSQQALGUR
SNAASTNDGSSRGIVSSISPHARVLAEGVCTLLNARVLPPAGAICFTSYDYM
DAYVLESTGYTAQISDVKRIKFKETRNGGGGGGGGRTAGTGGGGGAEIAEL
LREYQWISGEPASGAAERAPGSTTASSPVSTAASQOROKPSRRGAFPLVAMGG
RLSEINFADLGRVAVLGNVPYANPTDVERMNLKHIVATRLANADARBMGRTA
GSASTPSASSPFTCAEGLWYMDSMRVNQCIGRCIRHAGDYATILLDARTER
PDARRRVSNLQSMRVQAQTGQCFSGVREFFAERQPKG"
10078..11295
/gene="PAMT"
/note="L6910.10; predicted using Glimmer, Testcode and
CodonUsage"
10078..11295
/gene="PAMT"
/codon_start=1
/product="Protein Arginine Methyltransferase"
/protein_id="AAF35932.1"
/db_xref="GI:7025834"
/translation="MPANRMRFGAPTORGESRSANASKVLTESLATDAENKPTSK
SLYGDAAARLSNNIMTHDQRLURAYEYIFRNVRKTLHLGCGMGILYTLAARGMAK
MVTGVSSALVDAARVAEQNGKLNILIRCLVDVHLQPGMKFYDLCWGMPLLL
LNFRVLTALYAAHRLTPSGALCPNRAHLVAVSDYSFRLEDTEFWSNVYFGOMP
MKELVREYEMCAIPGNSIVSAPCLATHMTDLEGLTVEETATYEAQAOAASRON
EENPVHRVPTAVAGQGYEAATLSITRNATVHYLTFFVLDATFSTKTPGNFVLA
VRGGENNTEVSVGLREPLVPVNGEIQGLTIRIYTPADKGGKITVVEVTAKGVQA
AVETSGYIYQSY"
12821..13870
/gene="L6910.11"
/note="predicted using Glimmer, Testcode and CodonUsage"
12821..13870
/gene="L6910.11"
/codon_start=1
/product="L6910.11"
/protein_id="AAF35933.1"
/db_xref="GI:7025835"
/translation="MMSRVRPHIQRISVEDIMKGNRRLAKAITLAESNPADSTR
LGNLRIRHQNATVIRVFPALSGPGSGKSSLDLTUGYLCEKKNHVGVLAVDPS
TITHGSLGTQCKRKLHNSYIRPSPGHLGGVTSRAWEALETFEAREDDVVF
ETVGVGQSTQCKDLTDMILLVPPASGDELQGIKGVWEVADVVMVTKNGRRKPLV

Query Match 71.5%; Score 18.6; DB 3; Length 41944;
Best Local Similarity 84.0%; Pred. No. 6.7e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 cccgtctaggcggttcggtccgc 25

```

||||||| !! ||||||| |||||||
Db 38425 CCCGTCGGGGTTCGGCTCCGC 38449

RESULT 7
CFU96127/c 1432 bp mRNA linear MAM 28-JAN-1999
DEFINITION Canis familiaris adrenomedullin precursor, mRNA, complete cds.
ACCESSION U96127
VERSION U96127.1 GI:4193235
KEYWORDS
SOURCE dog.
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE 1 (bases 1 to 1432)
AUTHORS Ono,Y., Kojima,M., Okada,K. and Kangawa,K.
TITLE cDNA cloning of canine adrenomedullin and its gene expression in
the heart and blood vessels in endotoxin shock
JOURNAL Shock 10 (4), 243-247 (1998)
MEDLINE 99002704
REFERENCE 2 (bases 1 to 1432)
AUTHORS Ono,Y., Kojima,M., Okada,K. and Kangawa,K.
TITLE Direct Submission
JOURNAL Submitted (02-APR-1997) Biochemistry, National Cardiovascular
Center, Fujishirodai 5-7-1, Suita, Osaka 565, Japan
LOCATION/Qualifiers
FEATURES
source 1. .1432
/organism="Canis familiaris"
/db_xref="taxon:9615"
93. .155
sig_peptide /codon_start=1
CDS /product="adrenomedullin precursor"
93. .659 /protein_id="AAD09957.1"
/db_xref="GI:4193236"
/translation="MKLVFVALLYGLSLAFGLGADTARLDVASEFRKKKHWVSRGKR
ELRVSSSYPTGLAEVKGAPQATLIRTDQVKGASRNPTSGPDAARIRVKRYRQSMNMF
QGRPSFGCRFGTCTVOKLAHQIYQFTDNDKDGAVPRSKISPGQYGRRRRSLPEPGLR
RTLFLPEPRPGGAPAPRAHQVLANLLKM"
misc_feature 156. .215
/note="encodes proadrenomedullin N-terminal 20 peptide"
mat_peptide 375. .530
BASE COUNT 306 a 426 c 401 g 299 t
ORIGIN
Query Match 70.0%; Score 18.2; DB 4; Length 1432;
Best Local Similarity 87.0%; Pred. No. 1.1e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cccgtcaggcttcggttcgcg 23
| ||||| ||| ||||| |||||
Db 170 CACGCTAGCGGTGCGGTGCCG 148

RESULT 8
AF045773/c 1466 bp mRNA linear MAM 20-JAN-1999
LOCUS AF045773
DEFINITION Canis familiaris adrenomedullin precursor, mRNA, complete cds.
ACCESSION AF045773
VERSION AF045773.1 GI:4164449
KEYWORDS
SOURCE dog.
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE 1 (bases 1 to 1466)
AUTHORS Jougasaki,M., Schirger,J.A., Simari,R.D. and Burnett,J.C. Jr.
TITLE Autocrine role for the endothelin-B receptor in the secretion of
adrenomedullin
JOURNAL Hypertension 32 (5), 917-922 (1998)

MEDLINE 99040162
REFERENCE 2 (bases 1 to 1466)
AUTHORS Imoto,I. and Jougasaki,M.
TITLE Cloning of cDNA encoding canine adrenomedullin
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 1466)
AUTHORS Imoto,I. and Jougasaki,M.
TITLE Direct Submission
JOURNAL Submitted (02-FEB-1998) Molecular Cytogenetics, Mayo Clinic and
Foundation, 200 First Street, SW, Rochester, MN 55905, USA
REFERENCE 4 (bases 1 to 1466)
AUTHORS Imoto,I. and Jougasaki,M.
TITLE Direct Submission
JOURNAL Submitted (19-JAN-1999) Molecular Cytogenetics, Mayo Clinic and
Foundation, 200 First Street, SW, Rochester, MN 55905, USA
REMARK Sequence update by submitter
COMMENT On Jan 20, 1999 this sequence version replaced gi:3372495.
FEATURES
source 1. .1466
/organism="Canis familiaris"
/db_xref="taxon:9615"
148. .714
CDS /codon_start=1
/product="adrenomedullin precursor"
/db_xref="GI:4164450"
/translation="MKLVFVALLYGLSLAFGLGADTARLDVASEFRKKKHWVSRGKR
ELRVSSSYPTGLAEVKGAPQATLIRTDQVKGASRNPTSGPDAARIRVKRYRQSMNMF
QGRPSFGCRFGTCTVOKLAHQIYQFTDNDKDGAVPRSKISPGQYGRRRRSLPEPGLR
RTLFLPEPRPGGAPAPRAHQVLANLLKM"
BASE COUNT 308 a 441 c 411 g 306 t
ORIGIN
Query Match 70.0%; Score 18.2; DB 4; Length 1466;
Best Local Similarity 87.0%; Pred. No. 1.1e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cccgtcaggcttcggttcgcg 23
| ||||| ||| ||||| |||||
Db 225 CACGCTAGCGGTGCGGTGCCG 203

RESULT 9
AF147704
LOCUS AF147704
DEFINITION Streptomyces fradiae NDP-hexose 2,3-enoyl reductase TylCII
(tylCII), NDP-hexose 4-ketoreductase TylCIV (tylCIV), NDP-hexose
3-C-methyltransferase TylCIII (tylCIII), mycarosyl transferase
TylCV (tylCV), and NDP-hexose 3,5- (or5-) epimerase TylCVII
(tylCVII) genes, complete cds.
ACCESSION AF147704
VERSION AF147704.1 GI:5305790
KEYWORDS
SOURCE Streptomyces fradiae.
ORGANISM Streptomyces fradiae
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 5908)
AUTHORS Bate,N., Butler,A.R., Smith,I.P. and Cundliffe,E.
TITLE The mycarose-biosynthetic genes of Streptomyces fradiae, producer
of tylosin
JOURNAL Microbiology 146 (Pt 1), 139-146 (2000)
MEDLINE 10658660
PUBMED 10658660
REFERENCE 2 (bases 1 to 5908)
AUTHORS Bate,N. and Cundliffe,E.
TITLE Direct Submission
JOURNAL Submitted (30-APR-1999) Biochemistry, University Of Leicester,
University Road, Leicester LE1 7RH, UK
LOCATION/Qualifiers
FEATURES
source 1. .5908
/organism="Streptomyces fradiae"

```

```
/strain="T59235"
/db_xref="taxon.1906"
complement(80..1069)
/genes="tylCII"
/notes="orf10*"
complement(80..1069)
/genes="tylCII"
/function="biosynthesis of NDP-mycarose during tylosin
production"
/codon_start=1
/transl_table=11
/product="NDP-hexose 2,3-enoyl reductase TylCII"
/protein_id="AAD41821.1"
/db_xref="GI:5305791"
/translation="MSGMYVOLGRGATVLSRLWGTNFGSRVDPDDQALRLMDEALDR
GNCVTDADYGRWLYKGTTELVLGRLSGRRDDVLAIVGPEMSRVDNRGLSA
RHVIRCEASRLRGYDIDLQMHMRDTRVWDELQAMQDLVAGSKVRYIGSSNFA
GWLHAGQSAARGLSGVLYSEOCYLNLAHAELEVLPAARAYGIVGPAWPLHGLL
LSGALRLKLAEGTAVKSGOGRQRTLPALRDTIARYERFCARVGRDPAEYGLAWLISRP
GVSGAVIGPTTGHLYSALRAVELELSEEHRELEALFPVSGSGGEVPEAWQN"
complement(1066..2109)
/genes="tylCIV"
/notes="orf9*"
complement(1066..2109)
/genes="tylCIV"
/function="biosynthesis of NDP-mycarose during tylosin
production"
/codon_start=1
/transl_table=11
/product="NDP-hexose 4-ketoreductase TylCIV"
/protein_id="AAD41822.1"
/db_xref="GI:5305792"
/translation="MSGMLVPGVGRDRRLVLLGASGFLGSAAVAELARRPVRLRL
VARRSPVAGAAVEIEVRRDLSLPGVEAVEADAVVHLVATGEGKSKRAAGER
SEHNLVGLRELAEALRPTGPADATRAPVPLGSLQACMEQAHPGTYAOKLAAER
VLRHDAEGARVGLRLTVIGRSPLTSPGSGVIAVTAGRALAGDPITMWHDSVE
RLDLDVDAATFTALEHADQLRGKHWVGTGRRHLDRDVFETVAALAAERTGRPPV
PVYVDPGVAEVCDFRTPSDPSAFRAVTVGRPAEPADGIAAATAAVAGADSGPEP
EGAGKRGGGTERR"
complement(2106..3362)
/genes="tylCIII"
/notes="orf8*"
complement(2106..3362)
/genes="tylCIII"
/function="biosynthesis of NDP-mycarose during tylosin
production"
/codon_start=1
/transl_table=11
/product="NDP-hexose 3-C-methyltransferase TylCIII"
/protein_id="AAD41823.1"
/db_xref="GI:5305793"
/translation="MPAVPREDOMIISACVCGNRELLPVLDLGEOALTVGVPRTREE
TVPSIFLELVKSPGCGVLQVLRHTPDGLMYGEGVSGIRPFWIGHLRKVAAIR
EVLVDLDDLDIGSDATLLKAYPEGRLVIDPSGDFRELTPPHAEILAEYFSR
DVTAFRGTRARVITSIAMFYDLDPDLAFMRDHDVLDADGIWNEOSYLPAMLEAD
ADIVCHEHLEYALQOIEWMAERAGLTVLRAELTDVGGSLCVTLARASSPHRDEA
GPARIARETEAKINTMAPFEFARFVEHQDARDFLDRSAAAGRLTGYCASKGN
VILQYCGIGERDLPICIGVSPKAGRTFTGTGIPIVSEEDAKMRPDLQLLPLWYRE
GFVERERDPLAGGRLVFLPLRDVV"
3575..4960
/genes="tylCV"
/notes="orf7*"
3575..4960
/genes="tylCV"
/function="catalyses the incorporation of mycarose during
tylosin production"
/notes="glycosyltransferase"
/codon_start=1
/transl_table=11
/product="mycarosyl transferase TylCV"
/protein_id="AAD41824.1"
/db_xref="GI:5305794"
/translation="MAGLRPGAGVPEPTWPISPGKHCVRSISRRLPYHAPGTP
```

```
GRFRRADEGSCRMHAIAFFTLPAAGHVNPGLTGVAAELAAARGHRTVYALPEDMADRAV
RVGARAVTYPLDRERADMPKESDEYDEGEFLKVLWLLDTTADTLPLESFAE
EDRDVYANDPSTFTWGLLAGKWDIPVIRSTPSYASNEHVALHPPEPGAQVDPAL
IELTRAELKLEHGETSDPVAFAATVOSGPLEFYMPRIFOYAGETDDRRHFGPCAL
PRASFHTQWRQPEDGRPLVMSVLSGLTIYNERGIFRACVAFRDPWMLILVLGGGLGA
GDLGPLPENVLDRDFVLGDLVPLHTDILLVNHGGTSTAMEALAHGVPITVAMPPEPRA
TARRIAELDLGLWLLPGEVTAELSGIAQAVLTDDTRIRKGLDRMRGEIRRAGGPAVAA
DVIEGLLSPAA"
5004..5597
/genes="tylCVII"
/notes="orf6*"
5004..5597
/genes="tylCVII"
/function="biosynthesis of NDP-mycarose during tylosin
production"
/codon_start=1
/transl_table=11
/product="NDP-hexose 3,5- (or5-) eplmerase TylCVII"
/protein_id="AAD41825.1"
/db_xref="GI:5305795"
/translation="MIIETRVRDAYRITPEPIPDHRSLSVESLRVETLRATGHAIE
IRQVNHVNRNTLRLGHGTVPVPGQKIVTCVGAARTWVVDLVRVSGPTFGHDVVG
QDAESGVAVILPDGLGIVLADDDTCMNYLTITREYTPGMTIIDIDALDPLGLPWNLT
EPPVRSERDAAPSLAEAAAGTLPGEVQCLRAYPAM"
853 a 2123 c 2078 g 854 t
BASE COUNT
ORIGIN

Query Match 70.0%; Score 18.2; DB 1; Length 5908;
Best Local Similarity 87.0%; Pred. NO. 1.1e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 gcttagcgcttcggtcgccg 26
|||||
Db 4514 GTCGAGCGCTCCGTACCGGCC 4536
|||||

RESULT 10
SCF43
LOCUS Streptomyces coelicolor cosmid F43. linear BCT 17-JAN-2000
ACCESSION ALI36502
VERSION ALI36502.1 GI:6714724
KEYWORDS ABC transporter permease; ABC transporter, ATP-binding component;
bldA regulation; cytochrome P450; hydrolase; integral membrane
protein; LacI-family transcriptional regulatory protein; prolyl
aminopeptidase; RNA polymerase sigma factor; ROK (Nagc/YxLR) family
transcriptional regulator; secreted solute-binding lipoprotein;
Tetr-family transcriptional regulatory protein; TTA codon.
Streptomyces coelicolor A3(2).
Streptomyces coelicolor A3(2).
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 22396)
Redenbach,M., Kieser,H.M., Denapalte,D., Eichner,A., Cullum,J.,
Kinashi,H. and Hopwood,D.A.
A set of ordered cosmids and a detailed genetic and physical map
for the 8 Mb Streptomyces coelicolor A3(2) chromosome
Mol. Microbiol. 21 (1), 77-96 (1996)
97000351
MEDLINE 2 (bases 1 to 22396)
REFERENCE Saunders,D.C. and Harris,D.
AUTHORS Unpublished
JOURNAL
REFERENCE 3 (bases 1 to 22396)
AUTHORS Thomson,N.R., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
TITLE Direct Submission
JOURNAL Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
Submitted (17-JAN-2000) Streptomyces coelicolor sequencing project,
CB10 ISA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK
Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded
```

by the BBSRC and Beowulf Genomics
 Details of S. coelicolor sequencing at the Sanger Centre are
 available on the World Wide Web.
 (URL: http://www.sanger.ac.uk/Projects/S_coelicolor/)
 CDS are numbered using the following system eg SC7B7.01c. SC (S.
 coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
 strand).
 The more significant matches with motifs in the PROSITE database
 are also included but some of these may be fortuitous.
 The length in codons is given for each CDS.
 Usually the highest scoring match found by fasta -o is given for
 CDS which show significant similarity to other CDS in the database.
 The position of possible ribosome binding site sequences are given
 where these have been used to deduce the initiation codon.
 Gene prediction is based on positional base preference in codons
 using a specially developed Hidden Markov Model (Krogh et al.,
 Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot
 program of Bibb et al., Gene 30:157-66(1984) as implemented at
<http://www.nih.gov/jp/>
<http://cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the
 correct initiation codon. Where possible we choose an initiation
 codon (atg, gtg, ttg or (att)) which is preceded by an upstream
 ribosome binding site sequence (optimally 5-13bp before the
 initiation codon). If this cannot be identified we choose the most
 upstream initiation codon.
 IMPORTANT: This sequence MAY NOT be the entire insert of the
 sequenced clone. It may be shorter because we only sequence
 overlapping sections once, or longer, because we arrange for a
 small overlap between neighbouring submissions.
 Cosmid F43.

FEATURES

Location/Qualifiers
 1. .22396
 /organism="Streptomyces coelicolor A3(2)"
 /strain="A3(2)"
 /db_xref="taxon:100226"
 /clone="cosmid F43"
 1. .298
 /gene="SCF43.01"
 <1. .298
 /note="SCF43.01"
 /note="SCF43.01, possible hydrolase, partial CDS, len: >
 98 aa. Identical over the available sequence to
 Streptomyces lividans TR:P72465(EMBL:Z70724) substrate
 hyphae associated protein, Orf2 (218 aa), fasta scores
 opt: 648 z-score: 826.2 E(): 0.100.0 identity in 98 aa
 overlap. Also similar to several phosphatases e.g.
 Streptomyces collinus TR:O9X654(EMBL:AF131877) Naph
 phosphatase (231 aa), fasta scores opt: 186 z-score: 245.7
 E(): 2.6e-06 41.9% identity in 86 aa overlap."
 /codon_start=2
 /transl_table=11
 /product="putative hydrolase."
 /protein_id="CAB66190.1"
 /db_xref="GI:6714725"
 /translation="IDADDALEATASADVDACKPAPPEVHALEAGVPAERAVFG
 DTWMDRAGSRVRCVCGGLGRADLEAGAEAVADPADLLASRLSPLA"
 462. .1601
 /gene="SCF43.02"
 462. .1601
 /note="SCF43.02"
 /note="SCF43.02, possible integral membrane protein, len:
 379 aa. Almost identical to Streptomyces lividans
 TR:P72466(EMBL:Z70724) substrate hyphae associated
 protein, Orf3 (329 aa), fasta scores opt: 2115 z-score:
 2334.5 E(): 0.99.4% identity in 329 aa overlap. Also weakly
 similar to Mycobacterium tuberculosis
 SW:Y08U_MYCTU(EMBL:Z77724) hypothetical 36.8 kD protein
 (355 aa), fasta scores opt: 208 z-score: 236.0 E():
 9.3e-06 26.1% identity in 303 aa overlap and to a
 downstream CDS, SCF43.07, putative integral membrane
 protein (419 aa), fasta scores opt: 276 z-score: 252.1
 E(): 1.2e-08 28.1% identity in 366 aa overlap. Note
 SCF43.02 is 50 aa longer than Streptomyces lividans Orf3.

Contains possible membrane spanning hydrophobic domains."
 /codon_start=1
 /transl_table=11
 /product="putative integral membrane protein."
 /protein_id="CAB66191.1"
 /db_xref="GI:6714726"
 /translation="MWEACAVGRAVRAWRGPRERDLVQSLKAAGAALIAWTAVG
 VWLGDPMALMAWVALVLYQATVYSSYRQAGQFRAICTGALLASAAQAINDDNTGAL
 VLSLPVLMVWNSRFGGGIYAATVAVSLGATVSAAGVGHGNGAALGAVIGVA
 VNALVFPFHLRDVRENLAAALRAGDLHVAADLRETQWDAAGWGTSDAARLER
 RLALYSARSWRSRLTSGRLRLRHAPTVVPEDEDRWSRVGTNGLATFTLAV
 AAEHTPPAPGPPVLDLYARLLGLDACHTEAGRWAGERPCADPATATETMEELR
 RLQEGLRHAGQGAARTAVLGLLQAEINLWAEIVPEPRESAQ"
 1887. .1891
 /gene="SCF43.03"
 1898. .2929
 /note="SCF43.03"
 /note="SCF43.03, unknown, len: 343 aa. Almost identical to
 Streptomyces lividans TR:P72467(EMBL:Z70724) substrate
 hyphae associated protein, Orf4 (343 aa), fasta scores
 opt: 2274 z-score: 2584.7 E(): 0.99.1% identity in 343 aa
 overlap. Also weakly similar to Streptomyces coelicolor
 TR:O9WX08(EMBL:AL079345) putative epimerase (353 aa),
 fasta scores opt: 275 z-score: 318.0 E(): 2.5e-10 25.9%
 identity in 340 aa overlap."
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein SCF43.03."
 /protein_id="CAB66192.1"
 /db_xref="GI:6714727"
 /translation="MSSATGSKIYVGTGATNGVTSVVRLLSEDPVGVVLGIARRIPD
 WSPAKTEAWALDLASEQDLSHFADADAVVHLAWAFQPHDPAITWRINVLGRIFP
 EAVAAGVPAALVHASSGVAYSPGKNHVADESMPHTGWPDAAVYCKEAYLERALDTFE
 RDHGIRVVRMPAFLEKRESAQRIFRGFLPGPAARPELLPFLPDVPLGVQAL
 HTDDARAYRLVRSADARGANLAAEPVDAELLGELLGVPRVRLPRAAARSIAAA
 WGMRLLPASPHLEDAVLRPLVMDCTRAVELGNGATRTATATEVLEFLRQAGAGAT
 EPMGRKRGV"
 complement(2947. .3255)
 /gene="SCF43.04c"
 complement(2947. .3255)
 /note="SCF43.04c, unknown, len: 102 aa."
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein SCF43.04c."
 /protein_id="CAB66193.1"
 /db_xref="GI:6714728"
 /translation="MSDRHREGPGENGAPVPRDMPDQAGDHDPEWAAASPTREQSEH
 RQGGPGERPVDDAESAGDAETGVPDVPDPTDEAGTGRQGNHSAHVPEHPVPDESS
 complement(3320. .4609)
 /gene="SCF43.05c"
 complement(3320. .4609)
 /note="SCF43.05c"
 /note="SCF43.05c, possible ROK (NagC/XylR) family
 transcriptional regulator, len: 429 aa. Similar to
 Anaerocellum thermophilum SW:XYLR_ANATH(EMBL:Z69782)
 xylose repressor (399 aa), fasta scores opt: 519
 z-score: 575.9 E(): 1.1e-24 28.1% identity in 402 aa
 overlap. Also highly similar to Streptomyces coelicolor
 TR:OR7855(EMBL:AF031013) putative transcriptional
 regulator SC8A6.21c (441 aa), fasta scores opt: 2034
 z-score: 2241.7 E(): 0.71.4% identity in 434 aa overlap.
 Contains a PAM match to entry PF00480 ROK, ROK family and
 a putative helix-turn-helix motif situated between
 residues 30. .51 (+3.09 SD)."
 /codon_start=1
 /transl_table=11
 /product="putative ROK (NagC/XylR) family transcriptional
 regulator."
 /protein_id="CAB66194.1"
 /db_xref="GI:6714729"

/translation="MAGRNGRTVRDLRRNRTAVLRQRYFDGSLRFLGPATGLSSG
SVSNVADIVADGLVEAGSDSDGRPTLLRVAPASGHMIGVDVGETRVVFLFDL
TLTALARERLAPORHDYDVLVGHVDCGAEVATAGIPPERLLGAGTGVPIVEHT
ADRGAVVHGQTIGWDVPLEALLRAGSPDPDVPCLIDNGAKTLGQAEWFGAGRGAR
NAVVLFGSGVSLVTPAEQGRVVEWGHLLTVRVGRRCGALGCLDEAYAGASLLI
ARWEGGRVPEGTDEETALTAMLAAYPADGAADPVALAVLETAETAYLGLASLLI
NLFPORILIGWAGLQGLRFLPVRHRAVSYALRHPARKVTVDLGRIGPDVAIVTGA
AILPLADFFAGRRRPEPAPVPPAWRRTALEERAPH"
complement(3785..4033)
/gene="SCF43.05c"
/note="Pfam match to entry PF00480 ROK, ROK family, score
22.80, E-value 1e-05"
complement(4614..4618)
4789..4793
4799..5293
/gene="SCF43.06"
4799..5293
/gene="SCF43.06"
/note="Scf43.06"
/note="SCF43.06, unknown, len: 164 aa. Highly similar to a
protein of unknown function from Streptomyces coelicolor

misc_feature

Query Match 70.0%; Score 18.2; DB 1; Length 22396;
Best Local Similarity 87.0%; Pred. No. 1e+03; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 cccgtctaggcgttcggtgtccg 23
||||| ||||||||| |||
Db 12511 CCCGCGAGCGTTCGCTGTCGG 12533

RESULT 11

AC098934/c
LOCUS Homo sapiens chromosome 1 clone RP11-480I12, WORKING DRAFT
DEFINITION
SEQUENCE, 6 unordered pieces.
ACCESSION AC098934 AL390963
VERSION AC098934.1 GI:16751904
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 173434)
Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
Direct Submission
Unpublished
2 (bases 1 to 173434)
Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
Direct Submission
Submitted (06-NOV-2001) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
On Nov 6, 2001 this sequence version replaced gi:13157584.

Center: University of Washington Genome Center
Center Code: UWGC
Web site: http://www.genome.washington.edu
Contact: uwgctgs@u.washington.edu
Drafting Center: SC
----- Project Information
Center project name: chr-1
Center clone name: RP11-480I12 (sc0719)
----- Summary Statistics
Sequencing vector: plasmid; L08752; 88% of reads
Chemistry: Dye-terminator ET; 89% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 170284 bases at least Q40
Consensus quality: 171568 bases at least Q30
Consensus quality: 172298 bases at least Q20
Insert size: 172934; sum-of-ontigs
Quality coverage: 8.7x in Q20 bases; sum-of-contigs

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 9114: contig of 9114 bp in length
* 9115 9214: gap of unknown length
* 9215 27205: contig of 17991 bp in length
* 27206 27305: gap of unknown length
* 27306 47651: contig of 20346 bp in length
* 47652 47751: gap of unknown length
* 47752 74121: contig of 26370 bp in length
* 74122 74221: gap of unknown length
* 74222 112038: contig of 37817 bp in length
* 112039 112139: gap of unknown length
* 112139 173434: contig of 61296 bp in length.
FEATURES
Location/Qualifiers
1. 173434
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone_lib="RP11-480I12"
/clone_lib="RPCI human BAC library 11"
1. 9114
/note="assembly_name:Contig112"
9215..27205
/note="assembly_name:Contig113"
27306..47651
/note="assembly_name:Contig114"
47752..74121
/note="assembly_name:Contig115"
74222..112038
/note="assembly_name:Contig117"
112139..173434
/note="assembly_name:Contig118"
BASE COUNT 47569 a 39082 c 38186 g 48026 t 571 others
ORIGIN

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

FEATURES	Gausstr. 20, 42097 Wuppertal, FRG	AF319543	Streptomyces venezuelae putative acetyl-CoA acetyltransferase and cystathionine beta-synthase (CbsSV) genes, complete cds.
source	Location/Qualifiers	LOCUS	
	1. .3149	DEFINITION	
	/organism="Streptomyces lincolnensis"	ACCESSION	
	/strain="78-11"	VERSION	
gene	/db_xref="taxon:1915"	AF319543.1	GI:12082814
	10. .432	KEYWORDS	
CDS	/gene="melC1"	SOURCE	
	10. .432	ORGANISM	
	/gene="melC1"		
	/codon_start=1		
	/transl_table=11		
	/product="Tyrosinase co-factor (MelC2)"		
	/protein_id="CAA64999.1"		
	/db_xref="GI:1200139"		
	/db_xref="SWISS-PROT:P55048"		
	/translation="MPRLRRRLTAALASGAGAGAGAAAPCARAHDHGPSDV		
	PFPCLLDLLELLDEVIKRGIOGPHRGHGHGTGYAVFLDGVELHVMRNAD		
gene	GWISVSHYDPVTPRAARAADVLOGAKLVPPAN"		
	465. .1286		
CDS	/gene="melC2"		
	465. .1286		
	/gene="melC2"		
	/codon_start=1		
	/transl_table=11		
	/product="Tyrosinase"		
	/protein_id="CAA65000.1"		
	/db_xref="GI:1200140"		
	/db_xref="SWISS-PROT:P55023"		
	/translation="WYRKQNLTADEKRRFVAVLSSSARVDYFVTHNEFIVAD		
	TNGERTGRSPFLPWHRRFLLEFALQSDVASALPYMDWSTDRASLSLWAPDF		
	LGSTGRNRVTDGPFRAATGVPITVRLDGRYLRALGGAGRELPTRAEYDVSLS		
	IFTYDAPWNASDGFNRHLEGRVNLHNRVHVWVGOMATGVSPNDPVVFLHAYI		
	DKLMAQWRHRTPAYVPAAGTPDVVDLDTMKPNHDSPADLLDHTGHVTFDT		
gene	1369. .1821		
CDS	/gene="ORF3"		
	1369. .1821		
	/gene="ORF3"		
	/codon_start=1		
	/transl_table=11		
	/product="putative RNaseH"		
	/protein_id="CAA65001.1"		
	/db_xref="GI:1200141"		
	/db_xref="SWISS-PROT:P55049"		
	/translation="MSKVESIEGVVHTATNQVTPETPPQMSGVERIEQRTDL		
	THWTSVNGVHKEFDAITEIQIDERVAVTTVGEAQGVVTFHRLDDDDHTKWLQM		
	DFHPSVTEKYGDKLGFVKRQTKDLEKKEIEERGQETGGWRGAVI"		
gene	complement(2204. .2557)		
CDS	/gene="rnh"		
	complement(2204. .2557)		
	/gene="rnh"		
	/codon_start=1		
	/transl_table=11		
	/product="putative RNaseH"		
	/protein_id="CAA65002.1"		
	/db_xref="GI:1200142"		
	/db_xref="SPTREMBL:Q54388"		
	/translation="MRERVAAACDGASKNPGPAGWVWVADDTFTTRWEAGALGKA		
	TNNVAELTALERLLAARDVPLEIRMDSQYAMKAVTTWLPQWRNGWKTAAAGKPVAN		
	QDLVVRIDELDGRS"		
BASE COUNT	480 a 1140 c 1071 g 458 t		
ORIGIN			
	69.2%; Score 18; DB 1; Length 3149;		
Query Match	80.8%; Pred. No. 1.4e+03;		
Best Local Similarity	0; Mismatches 5; Indels 0; Gaps 0;		
Matches 21; Conservative	0; Mismatches 5; Indels 0; Gaps 0;		
QY	1 cccgtctaggcgttcggtctcgccg 26		
DB	866 CCCGTCAGGCGCACGCGTGTGCGCC 841		
RESULT	13		
	Query Match 69.2%; Score 18; DB 1; Length 6927;		
Best Local Similarity	80.8%; Pred. No. 1.3e+03;		
Matches 21; Conservative	0; Mismatches 5; Indels 0; Gaps 0;		
QY	1 cccgtctaggcgttcggtctcgccg 26		
DB	866 CCCGTCAGGCGCACGCGTGTGCGCC 841		
RESULT	13		


```

transmembrane protein; ubiquitin activating enzyme.
Leishmania major.
SOURCE
ORGANISM
Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
REFERENCE
AUTHORS
1 (bases 1 to 40064)
Ivens, A.C., Lewis, S.M., Bagherzadeh, A., Zhang, L., Chan, H.M. and
Smith, D.F.
A physical map of the Leishmania major Friedlin genome
JOURNAL
MEDLINE
98146435
REFERENCE
AUTHORS
2 (bases 1 to 40064)
Klages, S., Borzym, K., Reinhardt, R., Beck, A., Ivens, A.C., Quail, M.,
Rajandream, M.A. and Barrell, B.G.
Direct Submission
JOURNAL
MEDLINE
98146435
TITLE
JOURNAL
MEDLINE
98146435
COMMENT
On Feb 2, 2000 this sequence version replaced gi:4760361.
see http://www.ebi.ac.uk/parasites/leish.html
Notes:
Details of leishmania sequencing at the Sanger Centre are available
on the World Wide Web.
see http://www.sanger.ac.uk/Projects/Lmajor/
CDS are numbered using the following system eg L6586.01. L6586
(cosmid name), .01 (first CDS)
To make the cosmid library Leishmania major Friedlin DNA was
partially digested with Sau3AI prior to cloning into BamHI site of
the cosmid shuttle vector cLHVG (Ryan et al. 1993)
131:145-150). The sequence of the packaged vector was determined by
Peter Myler and Ken Stuart at Seattle Biomedical Research
Institute, and is available as accession number U59231.
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous. The length
in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.
Gene prediction is done using:
(1)
The FramePlot program of Bibb et al.,
Gene 30:157-166(1984) as implemented
at http://www.nih.gov/jp/
jun/cgi-bin/frameplot.pl. (2)
codon preference based on the codon usage table for Leishmania at
http://www.kazusa.or.jp/codon/
(3)
the Hexamer program which was written by Richard Durbin as an
integral part of the ACEDB-based analysis tools for the C.elegans
Genome Sequencing Project. The program calculates the
log-likelihood score for a given DNA segment based on the frequency
of 6-mers, normalised for the base-pair composition of the genome.
The program was trained using a fasta file of confirmed Leishmania
major coding sequences (CDS), i.e. from ATG start codon to the stop
codon.
CAUTION: We may not have predicted the correct initiation codon.
Where possible we choose an initiation codon (atg) which is
preceded by a stretch of pyrimidines or part of a Kozak sequence.
If this cannot be identified we choose the most upstream initiation
codon. Transmembrane domains were predicted as implemented at the
TMHMM server: http://www.cbs.dtu.dk/services/TMHMM-1.0/
IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions. Cosmid L6586 is not
overlapped by any clones sequenced to date. It contains the left
end of PAC P460 (AL160762), and the right end of PAC P986
(AL390693).
FEATURES
Location/Qualifiers
1. 40064
/organism="Leishmania major"
/strain="Friedlin"

```

```

/db_xref="taxon:5664"
/chromosome="23"
/clone="cosmid L6586"
153..170
/notes="(gcgtgt)3"
488..512
/notes="poly-pyrimidine tract"
1170..1185
/notes="poly-pyrimidine tract"
complement(1314..1707)
/notes="region of BLASTN similarity to: AFI39112 Leishmania
peruviana clone AC16 microsatellite sequence., bases
1..394, 83% identity over 393 bases"
1337..3379
/genes="L6586.01"
1337..3379
/genes="L6586.01"
/notes="L6586.01, len = 679 aa, conserved hypothetical
protein; predicted pi = 9.8667; predicted coiled-coil
region at aa
590-610; good similarity to AF253317 Trypanosoma cruzi
LYTlp (LYTl) gene; contains two Pfam matches to entry
PF00400 WD40, WD domain, G-beta repeat; reasonable
similarity to Q9PS7, conserved hypothetical protein (554
aa, Neurospora crassa, EMBL: AL355932, CAB91444); Fasta
scores: E():0, 34.9% identity in 502 aa"
/label="L6586.01"
/product="conserved WD40 repeat domain protein"
/protein_id="CAB71230.1"
/db_xref="GI:6855407"
/db_xref="SP:PREMBL:Q9NF84"
/translation="MPPTSLTSSRSPPVATAAPRVGAQRKPSKEVQELIDKRTTEL
KQETGQHEKFIQGRVWDTAALNRGRTRRRLNRPAPKRAKVDMYIQEEDNIGL
LNDHEVADRSIQSDIASGVLDIQOKRYSLVDKLGPKYKIDFNSINGTHLLAGLRGH
MANTREVSFQEGTOLKDRISDAIFLDVHSMATAVAKKFKVMTKGTGTHLLSKMA
HMDRLGYLPHKMLLAATSTSTMOYLDISTGQVGTGKVPVNRDPTSCSLAVNPSNGV
AATCDLGGVYKFSPPVDFPLQLKGHGVEDICEHPNGRFLTLGGDHAKMYWDCR
TAEGVRCPEDFVIGHSRGTSLIIPSGEANPDFFYANPHEHRRKRVNTLL
DKLPDPTISMDIQVPGVNEKRLAENYENLRNLRKARAIKREKMRASKSLGGAATGL
LVGRDEVDDELYGKEAPKTEVKRQELERKMRKWKDKDSADKVRSKQTLASRI
VQRNRAQMDARNGVDFDENALDAETALAAASRAAKQHKKLARKEALEIANDIRGO
LRQVDYDLVPPAQRREKSEKSHGSDTAGRTGEGSSGPEAEARTNAALKRFLR"
1424..2232
/genes="L6586.01"
/notes="region of BLASTN similarity to: AF253317
Trypanosoma cruzi LYTlp (LYTl) gene, LYTl-a allele,
complete cds., bases 4054..4922, 66% identity over 868
bases"
complement(1578..2185)
/notes="region of BLASTN similarity to: AQ644878
RPC193-DpnII-25G1.TV RPC193-DpnII Trypanosoma brucei
genomic clone RPC193-DpnII-25G1, bases 1..608, 67%
identity over 607 bases"
2222..2332
/genes="L6586.01"
/notes="Pfam match to entry PF00400 WD40, WD domain, G-beta
repeat, score 25.60, E-value 0.0012"
2335..3112
/genes="L6586.01"
/notes="region of BLASTN similarity to: AF253317
Trypanosoma cruzi LYTlp (LYTl) gene, LYTl-a allele,
complete cds., bases 4974..5751, 71% identity over 777
bases"
2348..2449
/genes="L6586.01"
/notes="Pfam match to entry PF00400 WD40, WD domain, G-beta
repeat, score 7.00, E-value 66"
2630..3063
/genes="L6586.01"
/notes="region of BLASTN similarity to: AL483866 TA292D02P
Trypanosoma brucei TREU927 sheared genomic DNA Trypanosoma

```

```
brucei genomic clone 292d02 forward, bases 1. .434, 67%
identity over 433 bases"
repeat_region 3522. .3542
repeat_region /note="poly-pyrimidine tract"
repeat_region 3908. .3922
repeat_region /note="(ctc)5"
repeat_region 3908. .3926
repeat_region /note="poly-pyrimidine tract"
repeat_region 3995. .4010
repeat_region /note="poly-pyrimidine tract"
repeat_region complement(4052. .4089)
repeat_region /note="poly-pyrimidine tract"
repeat_region 4068. .4085
repeat_region /note="(ggggag)3"
repeat_region 4717. .4739
repeat_region /note="poly-pyrimidine tract"
repeat_region complement(5114. .5131)
repeat_region /note="poly-pyrimidine tract"
repeat_region complement(5359. .5376)
repeat_region /note="poly-pyrimidine tract"
repeat_region 5390. .5405
repeat_region /note="poly-pyrimidine tract"
repeat_region 5561. .5577
repeat_region /note="poly-pyrimidine tract"
misc_feature 5908. .6414
/note="region of BLASTN similarity to: A0847049
LMAJFV1_lm37e03.xl Leishmania major FV1 random genomic
library Leishmania major genomic clone LMAJFV1_lm37e03 3',
bases 1. .507, 100% identity over 506 bases"
misc_feature 6300. .6849
/note="hexamer gene prediction, score 72.0664; previously
part of L6586.02"
complement(6837. .7217)
/note="region of BLASTN similarity to: A0850952
LMAJFV1_lm37e03.yl Leishmania major FV1 random genomic
library Leishmania major genomic clone LMAJFV1_lm37e03 5',
bases 1. .183, 99% identity over 182 bases, bases 184. .380,
99% identity over 196 bases"
repeat_region 7815. .7826
repeat_region /note="(gct)4"
repeat_region 8687. .8698
repeat_region /note="(cctc)3"
repeat_region 8845. .8862
repeat_region /note="poly-pyrimidine tract"
repeat_region 8882. .8914
repeat_region /note="poly-pyrimidine tract"
repeat_region 9523. .9539
repeat_region /note="poly-pyrimidine tract"
repeat_region 9951. .9969

Query Match 69.2%; Score 18; DB 3; Length 40064;
Best Local Similarity 80.8%; Pred. No. 1.2e+03;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 ccggtctaggcgttcggtccggcc 26
||||| ||||| ||||| |||||
Db 25046 CCCGCTCTGGCATTCTGTCGCGACC 25021
```

Search completed: July 31, 2002, 14:01:53
Job time: 17555 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 12:20:35 ; Search time 165.21 Seconds
(without alignments)
107.049 Million cell updates/sec

Title: US-09-899-718A-4
Perfect score: 72
Sequence: 1 ccttcaggacgatgcttcg.....catgtgagcccaacagtgc 72

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCUS_COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	23.8	33.1	2529	US-08-883-795A-37	Sequence 37, Appl
2	23.2	32.2	87350	US-08-781-891-79	Sequence 79, Appl
3	23	31.9	3187	US-08-706-037-26	Sequence 1, Appl
4	23	31.9	3192	US-08-706-037-26	Sequence 26, Appl
5	23	31.9	3192	US-08-706-037-26	Sequence 26, Appl
6	22.6	31.4	2375	US-08-903-800A-5	Sequence 5, Appl
7	22.2	30.8	80161	US-09-036-987A-1	Sequence 1, Appl
8	22.2	30.8	80161	US-09-370-700-1	Sequence 1, Appl
9	22	30.6	1968	US-09-232-191-30	Sequence 30, Appl
10	22	30.6	1968	US-09-232-200-80	Sequence 80, Appl
11	22	30.6	1968	US-09-232-197-80	Sequence 80, Appl
12	22	30.6	1968	US-09-232-201-80	Sequence 80, Appl
13	22	30.6	246240	US-08-724-394A-20	Sequence 20, Appl
14	22	30.6	246240	US-08-724-394A-21	Sequence 21, Appl
15	22	30.6	246240	US-08-724-394A-22	Sequence 22, Appl
16	21.8	30.3	765	US-08-998-416-1046	Sequence 1046, Ap
17	21.8	30.3	876	US-09-315-793-61	Sequence 61, Appl
18	21.8	30.3	1279	US-08-985-950-5	Sequence 5, Appl
19	21.8	30.3	1728	US-08-985-950-7	Sequence 7, Appl
20	21.8	30.3	3057	US-08-551-459-3	Sequence 3, Appl
21	21.8	30.3	3466	US-08-551-459-5	Sequence 5, Appl
22	21.6	30.0	107	US-09-042-353-251	Sequence 251, App
23	21.6	30.0	107	US-08-758-417A-99	Sequence 99, Appl
24	21.6	30.0	108	US-08-053-131-134	Sequence 134, App
25	21.6	30.0	108	US-08-645-641-134	Sequence 134, App
26	21.6	30.0	108	US-07-853-408B-134	Sequence 134, App
27	21.6	30.0	108	US-08-096-762-134	Sequence 134, App

28	21.6	30.0	108	2	US-08-308-865-134	Sequence 134, App
29	21.6	30.0	108	5	PCT-US92-10983-134	Sequence 134, App
30	21.6	30.0	3278	1	US-07-778-890A-1	Sequence 1, Appl
31	21.4	29.7	2249	3	US-08-814-052-19	Sequence 19, Appl
32	21.4	29.7	2279	3	US-08-814-052-17	Sequence 17, Appl
33	21.4	29.7	2300	3	US-08-814-052-18	Sequence 18, Appl
34	21.4	29.7	3183	2	US-08-939-218A-1	Sequence 1, Appl
35	21.4	29.7	3192	1	US-08-940-661A-1	Sequence 1, Appl
36	21.4	29.7	3192	2	US-09-083-485-1	Sequence 1, Appl
37	21.2	29.4	1174	2	US-08-793-410-8	Sequence 8, Appl
38	21.2	29.4	1604	4	US-09-438-938-9	Sequence 9, Appl
39	21.2	29.4	4325	1	US-08-453-924-2	Sequence 2, Appl
40	21.2	29.4	4325	1	US-08-471-791-29	Sequence 29, Appl
41	21.2	29.4	4325	5	PCT-US91-01746-29	Sequence 29, Appl
42	21.2	29.4	9323	1	US-08-038-682-6	Sequence 6, Appl
43	21.2	29.4	9323	1	US-08-302-832-6	Sequence 6, Appl
44	21.2	29.4	9323	2	US-08-530-198-6	Sequence 6, Appl
45	21.2	29.4	9323	2	US-08-469-880-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-08-883-795A-37
Sequence 37, Application US/08883795A
Patent No. 5985607
GENERAL INFORMATION:
APPLICANT: Delcuv, Genevieve
INVENTOR: Awang, Gregor
TITLE OF INVENTION: Recombinant DNA Molecules and Expression
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: BERESKIN & PARR
STREET: 40 King Street West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5H 3Y2
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/883,795A
FILING DATE: 27-JUN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gravelle, Micheline
REGISTRATION NUMBER: 40,261
REFERENCE/DOCKET NUMBER: 7841-062
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 364-7311
TELEFAX: (416) 361-1398
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 2529 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: Rh 10
US-08-883-795A-37

Query Match 33.1%; Score 23.8; DB 2; Length 2529;
Best Local Similarity 59.7%; Pred. No. 2.3; Indels 0;
Matches 40; Conservative 0; Mismatches 27; Gaps 0;


```
;
; COUNTRY: Republic of Korea
; ZIP: 305-335
; ADDRESSEE: KIM, Chul-Ho
; STREET: Shindonga Apt. 2-1101, Yongjeon-dong, Dong-gu
; CITY: Daejeon
; STATE: Daejeon
; COUNTRY: Republic of Korea
; ZIP: 300-200
; ADDRESSEE: SOHN, Jung-Hoon
; STREET: Nuri Apt. 103-506, Wolpyung-dong, Seo-gu
; CITY: Daejeon
; STATE: Daejeon
; COUNTRY: Republic of Korea
; ZIP: 302-280
; ADDRESSEE: KANG, Hyun-Ah
; STREET: Kyungseong Keunmaul Apt. 125-1501, Kalma-dong,
; STATE: Seo-gu
; CITY: Daejeon
; STATE: Daejeon
; COUNTRY: Republic of Korea
; ZIP: 302-171
; ADDRESSEE: KIM, Hwa-Young
; STREET: #3-462, Bukahyun-3-dong, Seodaemun-gu
; CITY: Seoul
; STATE: Seoul
; COUNTRY: Republic of Korea
; ZIP: 120-190
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5inch, 1.44MB storage
; COMPUTER: IBM PC/AT
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/903,800A
; FILING DATE: 31-JUL-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: KR 97-3173
; FILING DATE: 31-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME:
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE:
; TELEFAX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2375 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; FEATURE:
; OTHER INFORMATION: glyceraldehyde-3-phosphate
; OTHER INFORMATION: dehydrogenase gene of Hansenula
; OTHER INFORMATION: polymorpha DL-1(ATCC 26012)
;
US-08-903-800A-5

Query Match 31.4%; Score 22.6; DB 2; Length 2375;
Best Local Similarity 64.2%; Pred. No. 6.8;
Matches 34; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

Qy 9 gacgatgttcggtcctaagacacaccttctgtctatgacatgtgagc 61
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1717 GACGAAGTGGGTCAACAAGATACCGGCTTGGCGTCAAGACAGAGATC 1665

RESULT 7
US-09-036-987A-1
; Sequence 1, Application US/09036987A
; Patent No. 6143526
```

```
;
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Merlo, Donald J.
; APPLICANT: Treadway, Patti J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; TITLE OF INVENTION: Production
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dow Agrosciences LLC Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/036,987A
; FILING DATE: 09-MAR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stuart, Donald R
; REGISTRATION NUMBER: 28,479
; REFERENCE/DOCKET NUMBER: 50,608
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317)337-4816
; TELEFAX: (317)337-4847
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 80161 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
;
US-09-036-987A-1

Query Match 30.8%; Score 22.2; DB 3; Length 80161;
Best Local Similarity 61.0%; Pred. No. 30;
Matches 36; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

Qy 4 ttacgacgatgcttcggtcctaagacacaccttctgtctatgacatgtgagcc 62
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 14307 TTGAGTGGGACGTCGGTGGCCCGCGGACCACTGGCGGAGATGGCCTGGAAGCC 14365

RESULT 8
US-09-370-700-1
; Sequence 1, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H
; APPLICANT: Broughton, Mary C
; APPLICANT: Crawford, Kathryn P
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Treadway, Patti J
; APPLICANT: Turner, Jan R
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 DIV1
; CURRENT APPLICATION NUMBER: US/09/370,700
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/36987
; EARLIER FILING DATE: 1998-03-09
; NUMBER OF SEQ ID NOS: 39
```


Search completed: July 31, 2002, 12:21:09
Job time: 11692 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 12:11:57 ; Search time 5855.71 Seconds
(without alignments)
165.954 Million cell updates/sec

Title: US-09-899-718A-4
Perfect score: 72
Sequence: 1 ccttcagacatgcttgc.....catgtgagcccaacagtgc 72

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_hic:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	27.2	37.8	450	12	AQ466003 HS-5139_B
C 2	27.2	37.5	544	10	BE750138 201331_MA
C 3	27.2	37.5	877	10	BF028674 601763910
C 4	26.8	37.2	340	12	BH103249
C 5	26.4	36.7	930	10	BG676499
C 6	26.4	36.7	2069	10	BG852601
C 7	26.4	36.1	305	12	BH176205
C 8	26.4	36.1	305	12	CNS07J0P
C 9	26.4	36.1	305	12	CNS07J0Q
C 10	26.4	36.1	447	12	AQ536358
C 11	26.4	36.1	557	9	AW836780
C 12	26.4	36.1	622	12	AQ320218
C 13	25.8	35.8	502	12	AQ716839
C 14	25.6	35.6	273	9	BB524309
C 15	25.6	35.6	625	10	BF143122
C 16	25.4	35.3	510	12	AQ785147
C 17	25.4	35.3	555	12	BH210823

18	25.4	35.3	581	12	A2941767
19	25.2	35.0	285	9	BA483511
20	25.2	35.0	413	12	AQ297755
21	25.2	35.0	471	12	AQ005736
22	25.2	35.0	482	12	A2230592
23	25.2	34.7	127	10	BE854968
24	25.2	34.7	193	10	BF183196
25	25.2	34.7	235	9	AV336150
26	25.2	34.7	253	9	BE140819
27	25.2	34.7	270	9	AI592234
28	25.2	34.7	394	9	AA169000
29	25.2	34.7	465	9	AU179961
30	25.2	34.7	518	12	AQ433466
31	25.2	34.7	576	10	BM253384
32	25.2	34.7	620	12	AQ722509
33	25.2	34.7	636	10	BF725277
34	25.2	34.7	673	9	BB322109
35	25.2	34.7	674	10	BQ202917
36	25.2	34.7	747	10	BF664154
37	25.2	34.7	769	10	BG484311
38	25.2	34.7	1161	10	BG331922
39	24.8	34.4	152	10	R56980
40	24.8	34.4	231	10	BF955906
41	24.8	34.4	406	10	BE524520
42	24.8	34.4	635	12	A2111630
43	24.8	34.4	642	12	A2951598
44	24.8	34.4	661	12	A2496954
45	24.8	34.4	663	12	BH211493

ALIGNMENTS

RESULT 1
AQ466003/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

AQ466003 450 bp DNA linear GSS 23-APR-1999
HS-5139_B2_E07_SP6E RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate=715 Col=14 Row=J, DNA sequence.
AQ466003 1 GI:4643098
GSS.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 450)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.hnsc.washington.edu
Plate: 715 row: J column: 14
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 450.
Location/Qualifiers
1. .450
/organism="Homo sapiens"

A2941767 2M0201H07
BB483511 BB483511
AQ297755 HS-3109_B
AQ005736 CTF-HSP-2
A2230592 RPCI-23-7
BE854968 ux35h10.Y
BF183196 601809589
AV336150 AV336150
BE140819 PMO-HT002
AI592234 mr31h02.Y
AA169000 mr31h02.Y
AU179961 AU179961
AQ433466 HS-5052_B
BM253384 514426_MA
AQ722509 HS-5228_A
BF725277 bx14a11.Y
BB322109 BB322109
BQ202917 BQ202917
BF664154 602145956
BG484311 602504944
BG331922 602433085
R56980 F0026 Fetal
BF955906 RC4-NN17
BE524520 M51A3STM
A2111630 RPCI-23-4
A2951598 2M0216A09
A2496954 1M0333A12
BH211493 Sml-54L5.

Query Match 37.5%; Score 27; DB 10; Length 544;

Query Match	37.5%	Score 27;	DB 10;	Length 877;
Best Local Similarity	62.7%;	Pred. No. 38;		
Matches 42;	Conservative 0;	Mismatches 25;	Indels 0;	Gaps 0;
2	ctttcaggagcatgtccttcggtgccttaagacaccctacctttgtgtctatgacatgtgagc	61		
781	CTTACTGATPACATGCTGCTTGTCGAGACCGCTACTTGTCTCTCCGCAATGTGT	722		
62	ccaacag	68		
721	CCCTCAG	715		

RESULT 4
 BHI03249
 LOCUS
 DEFINITION BHI03249 340 bp DNA linear GSS 19-JUL-2001
 RPCI-24-230C14.TJ RPCI-24 Mus musculus genomic clone RPCI-24-230C14
 , DNA sequence.
 ACCESSION BHI03249
 VERSION BHI03249.1 GI:14930007
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 340)
 Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
 Tsegaye,G., Geer,K., Krol,M., Shwartsbeyn,A., Gebregeorgis,E.,
 Russell,D., de Jong,P. and Fraser,C.M.
 Mouse BAC End Sequences from Library RPCI-24
 Unpublished (1999)
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@igr.org
 Clones are derived from the mouse BAC library RPCI-24. For BAC
 library availability, please contact Pieter de Jong
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC
 Resources (<http://www.choi.org/bacpac/orderingframe.html>). BAC end
 page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 230 row: C column: 14
 Seq primer: SP6
 Class: BAC ends.

FEATURES
 source
 1..340
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-24-230C14"
 /clone_lib="RPCI-24"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /note="Vector: pTARBAC1; Site-1: BamHI; Site-2: BamHI;
 RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
 library was cloned in the pTARBAC1 cloning vector at the
 BamHI sites using MboI partially digested male C57BL/6J
 DNA."

BASE COUNT 88 a 77 c 96 g 79 t
 ORIGIN
 Query Match 37.2%; Score 26.8; DB 12; Length 340;
 Best Local Similarity 68.5%; Pred. No. 34;
 Matches 37; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
 Qy 15 gcttcggtccttaagacacactcttggctgtatgacatggagcccaacag 68
 ||||| | ||||| | | ||||| | ||||| | ||||| | |||||
 Db 178 gcttctcttttaagacaagaactgtgtgtgttgaatgtgcgtcaagag 231

RESULT 5
 BG676499
 LOCUS
 DEFINITION 602622960F1 NCI_CGAP_Skn4 Homo sapiens cDNA clone IMAGE:4747838 5',
 mRNA sequence.
 ACCESSION BG676499
 VERSION BG676499.1 GI:13907896
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 930)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 JOURNAL
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaaps-r@mail.nih.gov
 Tissue Procurement: James Cleaver, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM10598 row: h column: 15
 High quality sequence stop: 842.
 Location/Qualifiers

FEATURES

source
 1..930
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4747838"
 /clone_lib="NCI_CGAP_Skn4"
 /tissue_type="squamous cell carcinoma"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: Skin; Vector: pCMV-SPORT6; Site-1: NotI;
 Site-2: SalI; Cloned unidirectionally. Primer: oligo dr.
 Average insert size 1.5kb. Library constructed by Life
 Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 195 a 272 c 276 g 187 t
 ORIGIN

Query Match 36.7%; Score 26.4; DB 10; Length 930;
 Best Local Similarity 65.0%; Pred. No. 63;
 Matches 39; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
 Qy 13 atgttcggtgccttaagacacactcttggctgtatgacatggagcccaacagtggc 72
 ||||| | ||||| | | ||||| | ||||| | ||||| | |||||
 Db 796 atgttcagtcgccttcacttccttctctctctctctctctctctctctctctct 855

RESULT 6
 BG852601/c
 LOCUS
 DEFINITION 1024034F08.xl C. reinhardtii CC-1690, normalized, Lambda Zap II
 Chlamydomonas reinhardtii cDNA, mRNA sequence.
 ACCESSION BG852601
 VERSION BG852601.1 GI:14233785
 KEYWORDS EST.
 SOURCE Chlamydomonas reinhardtii.
 ORGANISM Chlamydomonas reinhardtii
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 Chlamydomonadaceae; Chlamydomonas.

REFERENCE 1 (bases 1 to 2069)
 Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,
 McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
 Analyses of the Chlamydomonas reinhardtii Genome: A Model,
 Unicellular System for Analyzing Gene Function and Regulation in
 Vascular Plants; project phase 2
 Unpublished (2000)
 Contact: Charles Hauser
 DCMB Box 91000
 Duke University
 Durham, NC 27708-1000
 Tel: 919 613 8159
 Fax: 919 613 8177
 Email: chauser@duke.edu.
 Location/Qualifiers

FEATURES
 source
 1..2069
 /organism="Chlamydomonas reinhardtii"
 /strain="CC-1690 wild type mt+ 21gr"
 /db_xref="taxon:3055"
 /clone_lib="C. reinhardtii CC-1690, normalized, Lambda zap
 II"

2: /note=Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
3: /XhoI; This library, constructed by John Davies and Jeffrey
4: Mc Dermott, combines cDNAs from CC-1690 cells grown to
5: mid-log phase in TAP (acetate-containing) medium in the
6: light, TAP medium in the dark, HS (minimal) medium in
7: ambient levels of CO₂ and HS medium bubbled with 5% CO₂.
8: PolyA mRNA was purified from each sample, pooled and cDNA
9: synthesized. The cDNA was directionally cloned into lambda
10: ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
11: pBluescript II SK- plasmids were excised from the lambda
12: ZAP clones by superinfection with EXASist (Stratagene)
13: phage. The library was normalized using method 4 described
14: in Bonaldo et al (1996), Genome Research 6: 791-806."
15: 559 c 637 g 275 t 61 others

BASE COUNT	537 a	559 c	637 g	275 t	61 others
ORIGIN					
Query Match		36.7%	Score 26.4;	DB 10;	Length 2059;
Best Local Similarity		62.9%	Pred. No. 79;		
Matches 39; Conservative		0;	Mismatches 23;	Indels 0;	Gaps 0;
QY	11	cgatgcttcggcgcccttaagacacactaccttctgtctatcacatgtgagcccaacatg	70		
Db	2017	CGATGTCGCNGTCGTTGAGCCATCTCTTTATGTAANGACATCATCAGTAACGGTG	1958		
QY	71	gc	72		
		-			
Db	1957	AC	1956		

RESULT	7
BHI76205/c	
LOCUS	BHI76205
DEFINITION	Schistosoma mansoni genomic clone 005F08 5', DNA sequence.
ACCESSION	BH176205
VERSION	BH176205.1
KEYWORDS	GSS.
SOURCE	Schistosoma mansoni.
ORGANISM	Schistosoma mansoni. Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE	1 (bases 1 to 305)
AUTHORS	Le Paslier,M.-C., Pierce,R.J., Merlin,F., Hirai,H., Wu,W., Williams,D.L., Johnston,B., Loverde,P.T. and Le Paslier,D.
TITLE	Construction and characterization of a Schistosoma mansoni bacterial artificial chromosome library
JOURNAL	Genomics 65 (2), 87-94 (2000)
MEDLINE	20247247
COMMENT	Contact: Pierce RJ

Institut Pasteur de Lille
1 rue du Professeur A. Calmette, 59019-Lille, France
Tel: (33) (0)3 20877783
Fax: (33) (0)3 20877888
Email: Raymond.Pierce@pasteur-lille.fr
CNS sequencing ID=DGCA005DC04CP2 Bases 140-286 have 85% identity
to S.mansoni EST AI395792.1 from base 137-281. Bases 140-258 have
84% identity to S.mansoni EST AI395362.1 from base 137-253.
Place: 005 row: F column: 08
Seq primer: M13 -21 primer
Class: BAC ends
High quality sequence stop: 305.

```

Query Match      36.1%; Score 26; DB 12; Length 305;
Best Local Similarity 65.5%; Pred. No. 63;
Matches 38; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 1 ccttcagagacgatcttcggtgccttaagacacctaccttctgtatgacatg 58
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 131 CATCTCCTTACGATGCTCACTGCCTTATGGATCAGACCTTTGGTCAATGGCTCGG 74

RESULT 9
CNS07J00/c

```

BASE COUNT
60 a 78 c 86 q 76 t 5 others

BASE COUNT	60 a	78 c	86 g	76 t	5 others
ORIGIN					
Query Match		36.1%	Score 26;	DB 12;	Length 305;
Best Local Similarity		65.5%;	Pred. No. 63;		
Matches 38;	Conservative	0;	Mismatches 20;	Indels 0;	Gaps 0;
QY	1	ccctttcaggacgacgtcttcggtgccttaagaacaccttactttgtgtctatgacatgtg	58		
Db	131	CATCTCCTTACGATGTCCTCATGCTTTATGGATACAGACCTTTGGGTCAATGCTGGG	74		

RESULT	8
CNS07J0P/c	
LOCUS	
DEFINITION	CNS07J0P 305 bp DNA linear GSS 08-OCT-2001 T7 end of clone 005DC04 of library SmbAC1 from strain Puerto-Rican of Schistosoma mansoni, genomic survey sequence.
ACCESSION	AL613147
VERSION	AL613147.1 GI:16026353
KEYWORDS	GSS.
SOURCE	Schistosoma mansoni.
ORGANISM	Schistosoma mansoni. Eukaryota; Metazoa; Platyhelminthes; Digenea; Strigeida; Schistosomatoidae; Schistosomidae; Schistosoma. 1 (bases 1 to 305) Le Paslier,M.C., Pierce,R.J., Merlin,F., Hirai,H., Wu,W., Williams,D.L., Johnston,D., LoVerde,P.T. and Le Paslier,D. Construction and characterization of a Schistosoma mansoni bacterial artificial chromosome library Genomics 65 (2), 87-94 (2000)
TITLE	20247247
JOURNAL	2 (bases 1 to 305)
MEDLINE	Genoscope.
REFERENCE	Direct Submission
AUTHORS	
TITLE	Submitted (05-OCT-2001) Genoscope - Centre National de Sequencage ;

JOURNAL

Submitted (05-OCT-2001) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

COMMENT

Partially Hind III digested and size-selected S. mansoni cercarial
DNA was ligated into Hind III digested pBlueBAC 11 vector and used
to transform E. coli DH10B. The complete library contains 23808
clones from 4 independent sizing-ligation-transformations. Average
insert size ranges from 70-127 kb and genome coverage is 7.9-fold.

FEATURES

Location/Qualifiers

1..305

/organism="Schistosoma mansoni"
/strain="Puerto-Rican"
/db_xref="taxon:6183"
/clone="005DC04"
/clone_lib="SmbAC1"
/note="end : T7"

60 a 78 c 86 q 76 t 5 others

BASE COUNT

Fax: 301 838 0208
Email: hbeatig@org

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genet cs (<http://info@resgen.com>). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.

Seq primer: SP6
Class: BAC ends.

Location/Qualifiers
1. .447
/organism="Homo sapiens"
/db_xref="GDB:7622025"
/db_xref="taxon:9606"
/clone="RPCI-11-318M10"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="lymphocytes"
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI; RPC111 Human Male BAC Library"

BASE COUNT
141 a 91 c 100 g 115 t

Query Match 36.1%; Score 26; DB 12; Length 447;
Best Local Similarity 70.0%; Pred. No. 70;
Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 16 ctctcggtccttaagacacactaccttctgtctgtctatgacatgtgagcccaa 65
||| | | |||| | |||| | |||| | | |||| |||| |||| ||||
Db 310 CTTTCATCATCATAGTCATCTACATGAGTTGTGTCTTCACATGTGAGGTCAA 359

RESULT 11
AW836780 557 bp mRNA linear EST 18-MAY-2000
QV1-LT0036-150200-070-f01 LT0036 Homo sapiens cDNA, mRNA sequence.
AW836780
AW836780.1 GI:7930754
EST.
KEYWORDS
SOURCE human.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 557)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brundstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-QV1-LT0036-150200-070-f01&t3=2000-02-15&t4=1>)
Seq primer: puc 18 forward
High quality sequence start: 11
High quality sequence stop: 557.
Location/Qualifiers
1. .557

FEATURES
source

Thu Aug 1 08:30:01 2002

DEFINITION	BB524309 RIKEN full-length enriched, 15 days embryo head Mus musculus cDNA clone D930019D20 3', mRNA sequence.
ACCESSION	BB524309
VERSION	BB524309.1
KEYWORDS	EST
	GI:9575767
	BASE COUNT 64 a 75 c 63 g 71 t
	ORIGIN
	FLC I"
	was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda

SOURCE	ORGANISM	Query Match	Score	DB	Length
house mouse.	Mus musculus	35.6%	25.6	9	273
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1 (bases 1 to 273)	Best Local Similarity	59.7%	Pred. NO.	84
AUTHORS	Konno.H., Aizawa.K., Akahira.S., Akiyama.J., Arakawa.T., Carninci	Matches	43	Conservative	0
		Mismatches	29	Indels	0
		Gaps	0		0

	QY	DB	QY	DB	QY	DB
P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J.J., Kikuchi, N., Kiyoasa, H., Kojlma, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y.	1	98	61	38	72	27
ccttcaggacgacgatgccttcgggtgccttaagaacaccactaccttgcttcatcgcattgacatggaag						
CTTTCAGGAAAATTTCTCAGGTACTTTAGTCAGCAACCTCACCGTCGCTGCCTGACCTGGGAT						
cccaacacagtggc			CCAAATAACTGCC			

TITLE	RIKEN Mouse ESTs (Konno, H., et al.)		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Yoshihide Havashizaki		
	Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tominaga, N., Toya, T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamanaka, I., Yano, K., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.		
		RESULT 15	
		BF143122	
		LOCUS	BF143122
		DEFINITION	601788012F1 NC1_CGAP_Lu30 Mus musculus cdna clone IMAGE:4015898 5', mRNA linear
			EST 24-OCT-2000

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216	<p>ACCESSION BF143122</p> <p>VERSION BF143122.1</p> <p>KEYWORDS EST.</p> <p>SOURCE house mouse.</p> <p>ORGANISM Mus musculus</p> <p>Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;</p>
---	---

<p> Email: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/ Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Thermotabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2) 520-524 (1998) </p>	<p> REFERENCE 1 (bases 1 to 625) Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus </p>
--	---

Automated filtration-based high-throughput plasmid preparation system. *Genome Res.* 9 (5), 463-470 (1999)

Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. *Methods Enzymol.* 303, 3-14 (1999)

Itouh, M., Kitsuunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Tissue Procurement: Gilbert Smith, Ph.D.

19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES	source
Location/Qualifiers	
1..273	
/organism="Mus musculus"	
/strain="C57BL/6J"	
/db_xref="taxon:10090"	
/clone="D930019D20"	
/clone_lib="RIKEN full-length enriched, 15 days embryo	
1..625	
/organism="Mus musculus"	
/strain="CZECH II"	
/db_xref="taxon:10090"	
/clone="IMAGE:401898"	
/clone_lib="NCI_CGAP_Lu30"	
/tissue_type="tumor, metastatic to mammary"	
source	

```

head"
/sex="mixed"
/tissue_type="head"
/dev_stage="15 days embryo"
/lab_host="DH10B"
/note="Site.1: Sall; site.2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
"
BASE COUNT      140 a      130 c      122 g      233 t
Investigator providing samples: Gilbert Smith, NIH
dr. Library constructed by Life Technologies.
MMTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo
Site.2: Sall; transgenic model WNT-1, expression driven by
/note="Organ: lung; Vector: pCMV-SPORT6; Site.1: NotI;
/lab_host="DH10B"

```

Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGAGTCCTTTTTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse	ORIGIN	Query Match Best Local Similarity Matches 34; Conservative	35.6%; 70.8%; 0;	Score 25.6; DB 10; Pred. NO. 1.1e+02; Mismatches 14;	Length 625; Indels 0; Gaps 0;
---	--------	--	------------------------	--	----------------------------------

transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence 5',

GAGAGAGATTCGAGTTAATTAAATTAATCCCCCCCCCCC 3'. cDNA

	Qy	1	ccttcaggacgatcgtcggtgcttaagaccaccttcttgct	48
	Db	578	CCTTTTACCACAATCCTCTGTACTTAAGAACCTGAATAATGCT	625

Thu Aug 1 08:30:01 2002

us-09-899-718a-4.rst

Page 8

Search completed: July 31, 2002, 12:12:01
Job time: 13474 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 14:11:11 ; Search time 720.8 Seconds
(without alignments)
171.501 Million cell updates/sec

Title: us-09-899-718a-4
Perfect score: 72
Sequence: 1 ctttcagagcagctctg.....catgtgagcccaagtggtc 72

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25.6	35.6	3085	23	ABL16797
2	25.6	35.6	20261	23	ABL16796
3	25	34.7	152	21	AAC07947
4	25	34.7	424	21	AAC00215
5	25	34.7	626	20	AA13569
6	25	34.7	907	20	AA253005
7	25	34.7	3098	22	ABA18125
8	25	34.7	3098	22	AA540421
9	25	34.7	3098	22	AA104024

1	25.6	35.6	3085	23	ABL16797
2	25.6	35.6	20261	23	ABL16796
3	25	34.7	152	21	AAC07947
4	25	34.7	424	21	AAC00215
5	25	34.7	626	20	AA13569
6	25	34.7	907	20	AA253005
7	25	34.7	3098	22	ABA18125
8	25	34.7	3098	22	AA540421
9	25	34.7	3098	22	AA104024

c 10	25	34.7	3098	22	AAK86871	Human immune/haema
c 11	25	34.7	4316	22	ABA83122	H1A-Cw ovarian tum
c 12	24.6	34.2	50000	24	AAAD26400	Human glutamate re
c 13	24.6	34.2	50000	24	AAAD26437	Human GRM3 gene fr
c 14	24.2	33.6	479	22	ABA43910	Human foetal liver
c 15	24.2	33.6	479	22	ABA54363	Human breast cell
c 16	24.2	33.6	479	22	ABA24139	Probe #2605 for ge
c 17	24.2	33.6	479	22	AAK02657	Human brain expres
c 18	24.2	33.6	479	22	AAK28093	Human bone marrow
c 19	24.2	33.6	479	22	AAI12672	Probe #2605 for ge
c 20	24.2	33.6	479	22	AAI34021	Probe #2707 used t
c 21	24.2	33.6	479	22	AAI02580	Probe #2571 used t
c 22	24.2	33.6	15464	23	ABL06682	Drosophila melanog
c 23	24	33.3	388	22	AAI86498	Human polynucleoti
c 24	23.8	33.1	2529	17	AAI31531	Human 3' apolipop
c 25	23.8	33.1	2529	21	AAI31997	Human apolipoprote
c 26	23.4	32.5	273	22	AAI29187	Colon tumour relat
c 27	23.4	32.5	409	23	AAI39790	DNA encoding novel
c 28	23.4	32.5	435	22	AAI29916	Human lung antigen
c 29	23.4	32.5	747	23	AAI58532	cdNA #1208 encodin
c 30	23.4	32.5	2016	11	AAI02887	cdNA of human plac
c 31	23.4	32.5	2360	11	AAI02888	Human placenta-der
c 32	23.4	32.5	3300	21	AAI05446	Streptococcus pneu
c 33	23.4	32.5	6588	22	AAH62768	Shrimp white spot
c 34	23.4	32.5	15213	19	AAV52159	Streptococcus pneu
c 35	23.4	32.5	305107	22	AAH62689	Shrimp white spot
c 36	23.2	32.2	11628	20	AAH20520	Polynucleotide seq
c 37	23.2	32.2	87350	18	AAH83003	Human WRN genomic
c 38	23	31.9	3012	22	AAH98552	Human ESR-derived
c 39	23	31.9	3187	17	AAI10922	Laccase gene. Myc
c 40	23	31.9	3192	18	AAI63318	Myceliophthora the
c 41	23	31.9	3192	21	AAI24236	M. thermophila lac
c 42	23	31.9	6868	22	AAI16293	Human App-murine n
c 43	23	31.9	11036	20	AAI13401	Enterococcus faeca
c 44	23	31.9	57728	22	AAH87588	Human 9p11 chromos
c 45	22.8	31.7	1001	21	AAH51258	Human GSHS related

ALIGNMENTS

RESULT 1

ABL16797
ID ABL16797 standard; DNA; 3085 BP.
XX AC ABL16797;
XX AC ABL16797;
DT 26-MAR-2002 (first entry)
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 1864.
DE Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -
XX Claim 1; SEQ ID NO 1864; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 3085 BP; 861 A; 776 C; 752 G; 696 T; 0 other;
SQ

Query Match 35.6%; Score 25.6; DB 23; Length 3085;
Best Local Similarity 59.7%; Pred. No. 3.7;
Matches 43; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 1 cctttcaggacgatgcttcggtccttaagacacaccttctgtctatgacatgtgag 60
||||| || ||| | |||| | | || | || | || | ||
Db 1321 cctttcggaaatagcatgagcctgaggagttcaaccgctctccaagagctgacag 1380

QY 61 ccccaacagtggc 72
|||| | || | |
Db 1381 tccaaaagtggc 1392

RESULT 2
ABL16796
ID ABL16796 standard; DNA; 20261 BP.
XX
AC ABL16796;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 1861.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmacetical; gene; ds.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Claim 1; SEQ ID NO 1861; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 20261 BP; 5591 A; 4571 C; 4371 G; 5728 T; 0 other;
SQ

Query Match 35.6%; Score 25.6; DB 23; Length 20261;
Best Local Similarity 59.7%; Pred. No. 6.7;
Matches 43; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 1 cctttcaggacgatgcttcggtccttaagacacaccttctgtctatgacatgtgag 60
||||| || ||| | |||| | | || | || | || | ||
Db 17207 cctttcggaaatagcatgagcctgaggagttcaaccgctctccaagagctgacag 17266

QY 61 ccccaacagtggc 72
|||| | || | |
Db 17267 tccaaaagtggc 17278

RESULT 3
AAC07947/c
ID AAC07947 standard; cDNA; 152 BP.
XX
AC AAC07947;
XX
XX 06-OCT-2000 (first entry)
XX
XX Human secreted protein 5' EST, SEQ ID NO: 12022.
XX
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
XX Homo sapiens.
XX
XX EP1033401-A2.
XX
XX 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-0200610.
XX
XX 26-FEB-1999; 99US-0122487.
XX
XX (GEST) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
PI WPI; 2000-500381/45.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX Claim 1; SEQ ID 12022; 71pp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX identified within the present sequence. The 5' ESTs were prepared from
XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX sequences usually correspond mainly to the 3' untranslated region (UTR)
XX of the mRNA because they are often obtained from oligo-dT primed cDNA
XX libraries. Such ESTs are not well suited for isolating cDNA sequences
XX derived from the 5' ends of mRNAs and even in those cases where longer
XX cDNA sequences have been obtained, the full 5' UTR is rarely included.
XX 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
XX used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
XX in diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX They are used to obtain upstream regulatory sequences and to design
XX expression and secretion vectors.

DT 14-MAR-2000 (first entry)
XX Human prostate tumor cDNA library derived EST fragment #148.
DE Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;
XX treatment; ds.
KW Homo sapiens.
KW DE19820190-A1.
OS
PN
XX
XX 04-NOV-1999.
PD
XX 28-APR-1998; 98DE-1020190.
PF
XX 28-APR-1998; 98DE-1020190.
PR
XX (META-) METAGEN GES GENOMFORSCHUNG MBH.
PA Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
XX WPI; 1999-621386/54.
PI
XX New human nucleic acid sequences from pancreatic tumors, and related
XX proteins -
PT
XX Claim 2; Page 299-300; 502pp; German.
PS
XX This invention describes novel polypeptides and their encoding nucleic
XX acids derived from human pancreatic tumor tissue which have cytostatic
XX activity. The sequences are also useful in producing pharmaceutical
XX compositions for treatment of pancreatic tumors. AAZ52858-Z53014
XX represent expressed sequence tag (EST) fragments derived from a human
XX pancreatic tumor cDNA library and which encode the proteins represented
XX in AAZ73814-Y74252.
XX
SQ Sequence 907 BP; 152 A; 266 C; 323 G; 166 T; 0 other;

Query Match 34.7%; Score 25; DB 20; Length 907;
Best Local Similarity 69.4%; Pred. No. 4.3;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 6 caggacgatgcttcggtgccttaagacacaccttctgtctatgaca 54
||||| ||||||| || ||||| ||||| ||||| |||||
Db 246 CAGGAGGAGGCTTCGGGGCGCATGACCCCAACCTCGCTGTGGGAAA 198

RESULT 7
ABA18125/c
ID ABA18125 standard; DNA; 3098 BP.
XX AC
XX ABA18125;
XX
XX 23-JAN-2002 (first entry)
DT
XX Human nervous system related polynucleotide SEQ ID NO 10456.
DE
XX Human; neurotropic; neuroprotective; cytostatic; dermatological; virucide;
XX immunosuppressive; antineoplastic; anti-HIV; antibacterial; vulnery;
XX antiparkinsonian; antitickling; antianemic; antiarthritic; cancer;
XX antineumatic; hepatotropic; cerebroprotective; antineoplastic;
XX antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
XX antiparasitic; cardiac; immune disorder; cardiovascular disorder;
XX neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
OS Homo sapiens.
XX
XX WO200159063-A2.
PN
XX 16-AUG-2001.
PD
XX 17-JAN-2001; 2001WO-US01334.
PF

XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUL-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 26-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.

PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244517.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249219.
PR 17-NOV-2000; 2000US-0249220.
PR 17-NOV-2000; 2000US-0249221.
PR 17-NOV-2000; 2000US-0249222.
PR 17-NOV-2000; 2000US-0249223.
PR 17-NOV-2000; 2000US-0249224.
PR 17-NOV-2000; 2000US-0249225.
PR 17-NOV-2000; 2000US-0249226.
PR 17-NOV-2000; 2000US-0249227.
PR 17-NOV-2000; 2000US-0249228.
PR 17-NOV-2000; 2000US-0249229.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251858.
PR 08-DEC-2000; 2000US-0251859.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Barash SC, Ruben SM;

XX WPI; 2001-541565/60.

XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system

PT cancers and metastases -
XX Disclosure; SEQ ID NO 10456; 1701pp + Sequence Listing; English.
PS
XX
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (AB14678-AB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3098 BP; 629 A; 858 C; 936 G; 675 T; 0 other;

Query Match 34.7%; Score 25; DB 22; Length 3098;

Best Local Similarity 69.4%; Pred No. 6.4;

Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 6 caggacgatcttcggtgccttaagacacaccttctgtgtctatgaca 54

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 57 CAGGAGGAGGCTCGGGGGCCATGACCCCAACCTCCGGCTCTGGGAAA 9

RESULT 8

AAS40421/c

ID AAS40421 standard; DNA; 3098 BP.

XX AAS40421;

XX 17-DEC-2001 (first entry)

DE DNA encoding human prostate cancer antigen, Seq ID No 573.

XX Human; prostate cancer antigen; cytostatic; uropathic; diagnostic;
KW reproductive system; chromosomal marker; forensic; urinary disorder;
KW chronic nephritis; blood-related disorder; thrombosis; ds.

XX Homo sapiens.

OS WO200155316-A2.

PN 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01328.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.


```

PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465570/50.
XX
XX Isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition
XX
XX Disclosure; SEQ ID NO 6712; 1297pp + Sequence Listing; English.
XX
XX The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.
XX
XX Sequence 3098 BP; 629 A; 858 C; 936 G; 675 T; 0 other;
SQ
Query Match 34.7%; Score 25; DB 22; Length 3098;
Best Local Similarity 69.4%; Pred No. 6.4;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
QY 6 caggacgatcctcggtgccttaagacacacctttgtgtctatgaca 54
||||| ||||| || ||||| ||||| ||||| |||||
Db 57 CAGGAGGAGGCTTCGGGGGCCCATGACCAACCTCCGCGTCTGGGAAA 9
RESULT 10
AAK6871/c
ID AAK6871 standard; DNA; 3098 BP.
XX
XX AAK6871;
XX
XX 07-NOV-2001 (first entry)
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:41683.
DE
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
OS
XX WO200157182-A2.
PN
XX 09-AUG-2001.
PD

```

```

XX 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218390.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.

```

PR	29-SEP-2000;	2000US-0236369;
PR	29-SEP-2000;	2000US-0236370;
PR	02-OCT-2000;	2000US-0236802;
PR	02-OCT-2000;	2000US-0237037;
PR	02-OCT-2000;	2000US-0237038;
PR	02-OCT-2000;	2000US-0237039;
PR	02-OCT-2000;	2000US-0237040;
PR	13-OCT-2000;	2000US-0239935;
PR	13-OCT-2000;	2000US-0239937;
PR	20-OCT-2000;	2000US-0240960;
PR	20-OCT-2000;	2000US-0241221;
PR	20-OCT-2000;	2000US-0241785;
PR	20-OCT-2000;	2000US-0241786;
PR	20-OCT-2000;	2000US-0241787;
PR	20-OCT-2000;	2000US-0241808;
PR	20-OCT-2000;	2000US-0241809;
PR	20-OCT-2000;	2000US-0241826;
PR	01-NOV-2000;	2000US-0246171;
PR	08-NOV-2000;	2000US-0246474;
PR	08-NOV-2000;	2000US-0246475;
PR	08-NOV-2000;	2000US-0246476;
PR	08-NOV-2000;	2000US-0246477;
PR	08-NOV-2000;	2000US-0246478;
PR	08-NOV-2000;	2000US-0246523;
PR	08-NOV-2000;	2000US-0246524;
PR	08-NOV-2000;	2000US-0246526;
PR	08-NOV-2000;	2000US-0246527;
PR	08-NOV-2000;	2000US-0246528;
PR	08-NOV-2000;	2000US-0246532;
PR	08-NOV-2000;	2000US-0246609;
PR	08-NOV-2000;	2000US-0246610;
PR	08-NOV-2000;	2000US-0246611;
PR	08-NOV-2000;	2000US-0246612;
PR	17-NOV-2000;	2000US-0249211;
PR	17-NOV-2000;	2000US-0249212;
PR	17-NOV-2000;	2000US-0249213;
PR	17-NOV-2000;	2000US-0249214;
PR	17-NOV-2000;	2000US-0249215;
PR	17-NOV-2000;	2000US-0249216;
PR	17-NOV-2000;	2000US-0249217;
PR	17-NOV-2000;	2000US-0249219;
PR	17-NOV-2000;	2000US-0249220;
PR	17-NOV-2000;	2000US-0249221;
PR	17-NOV-2000;	2000US-0249222;
PR	17-NOV-2000;	2000US-0249244;
PR	17-NOV-2000;	2000US-0249245;
PR	17-NOV-2000;	2000US-0249246;
PR	17-NOV-2000;	2000US-0249264;
PR	17-NOV-2000;	2000US-0249265;
PR	17-NOV-2000;	2000US-0249297;
PR	17-NOV-2000;	2000US-0249299;
PR	17-NOV-2000;	2000US-0249300;
PR	01-DEC-2000;	2000US-0250160;
PR	01-DEC-2000;	2000US-0250391;
PR	05-DEC-2000;	2000US-0251030;
PR	05-DEC-2000;	2000US-0251988;
PR	05-DEC-2000;	2000US-0256719;
PR	06-DEC-2000;	2000US-0251479;
PR	08-DEC-2000;	2000US-0251856;
PR	08-DEC-2000;	2000US-0251868;
PR	08-DEC-2000;	2000US-0251869;
PR	08-DEC-2000;	2000US-0251989;
PR	08-DEC-2000;	2000US-0251990;
PR	11-DEC-2000;	2000US-0254097;
PR	05-JAN-2001;	2001US-02595678;

PT	Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT	useful for preventing, diagnosing and/or treating cancers and
PT	metastasis -
XX	
PPS	Disclosure; SEQ ID NO 41683; 3071pp + Sequence Listing; English.
XX	
CC	AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I).
CC	amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC	activity, and can be used in gene therapy and vaccine production. (I)
CC	proteins and polynucleotides may be used in the prevention, diagnosis and
CC	treatment of diseases associated with inappropriate (I) expression. For
CC	example, they may be used to treat disorders associated with decreased
CC	expression by rectifying mutations or deletions in a patient's genome
CC	that affect the activity of (I) by expressing inactive proteins or to
CC	supplement the patients own production of (I). Additionally, (I)
CC	polynucleotides may be used to produce the secreted (I), by inserting
CC	the nucleic acids into a host cell and culturing the cell to express the
CC	protein. (I) proteins and polynucleotides may be used to prevent,
CC	diagnose and treat immune/hematopoietic-related diseases, especially
CC	cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC	to AAK87694 represent human immune/hematopoietic antigen genomic
CC	sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC	represent sequences used in the exemplification of the present invention.
XX	
XX	Sequence 3098 BP: 629 A; 858 C; 936 G; 675 T; 0 other;

Query Match 34.7%; Score 25; DB 22; Length 3098;
Best Local Similarity 69.4%; Pred. No. 6.4;
Matches 34: Conservative 0; Mismatches 15; Indels 0; Caps 0;

Qy 6 caggacgatgcttcgggtgccttaagacacccaccttttgtctatgaca 54
||||| ||| ||||| ||| | ||| ||| ||||| ||| ||| ||| |||

Dh 57 CAGGAGGAGGTTCGGGGCGGCATGACCCCACTCCGCGTCTGGGAAA 9

RESULT	11	
ABA83122/c		
ID	ABA83122 standard; DNA; 4316 BP.	
XX		
XX	ABA83122;	
XX		
XX	08-FEB-2002 (first entry)	
DT		
XX		
DE	HLA-Cw ovarian tumour marker gene, SEQ ID NO:81.	
XX		
XX	Ovarian tumour marker gene; human; overexpression; upregulation;	
KW	epithelial tumour; cancer; diagnosis; prognosis; disease monitoring;	
KW	identification; serous cystadenoma; borderline serous tumour;	
KW	serous cystadenocarcinoma; mucinous cystadenocarcinoma;	
KW	mucinous cystadenoma; borderline mucinous tumour; endometrioid carcinoma;	
KW	undifferentiated carcinoma; clear cell adenocarcinoma; cystadenofibroma;	
KW	adenofibroma; Brenner tumour; serial analysis of gene expression; SAGE;	
KW	immune response pathway; cell proliferation regulation; protein folding;	
KW	membrane localised; secreted; therapeutic target; cytostatic;	
KW	gene therapy; vaccine; ds.	
XX		
XX	Homo sapiens.	
OS		
XX		
PN	WO200175177-A2.	
XX		
PD	11-OCT-2001.	
XX		
XX	03-APR-2001; 2001WO-US10947.	
PF		
XX		
XX	03-APR-2000; 2000US-194336P.	
PR		
XX	(USSH) US DEPT HEALTH & HUMAN SERVICES.	
XX		
PA		
XX		
XX		
PI	Morin PJ, Sherman-Baust CA, Pizer ES, Hough CD;	
XX		
XX	WPI; 2001-626450/72.	
DR	P-PSDB; ABB50296.	
DR		

Search completed: July 31, 2002, 14:11:17
Job time: 17374 sec

Db 72 CTTTCAGAGCAGCAATGCTCCTGAGACACAGACCTTGCTATCACTAGCTGTGG 13
QY 62 ccaacagt 70
Db 12 CCCTGAGAG 4

RESULT 15
ABA54363/C
ID ABA54363 standard; DNA; 479 BP.
XX
AC ABA54363;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #2668.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX

(MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -

XX Claim 1; SEQ ID NO 2668; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC foetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.

CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 479 BP; 126 A; 134 C; 82 G; 137 T; 0 other;

Query Match 33.6%; Score 24.2; DB 22; Length 479;
Best Local Similarity 59.4%; Pred. No. 7.4;
Matches 41; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 2 ctttcagacatgcttcgtggtccttaagacacacaccttctgtgtatgacatgtgagc 61
|||||
Db 72 CTTTCAGAGCAGCAATGCTCCTGAGACACAGACCTTGCTATCACTAGCTGTGG 13

QY 62 ccaacagt 70
Db 12 CCCTGAGAG 4

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 14:00:51 ; Search time 6034.22 seconds

(without alignments)
249.694 Million cell updates/sec

Title: US-09-899-718a-4

Perfect score: 72
1 ccttcagagacatgcttcg.....catgtgagccacacagtgc 72

Scoring table: IDENTITY_NTC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 segs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Genbank: 1: gb_da: 2: gb_hcg: 3: gb_in: 4: gb_cm: 5: gb_ov: 6: gb_pat: 7: gb_ph: 8: gb_pl: 9: gb_pr: 10: gb_ro: 11: gb_sts: 12: gb_sy: 13: gb_un: 14: gb_vi: 15: em_ba: 16: em_fun: 17: em_hum: 18: em_in: 19: em_mu: 20: em_om: 21: em_or: 22: em_ov: 23: em_pat: 24: em_ph: 25: em_pl: 26: em_po: 27: em_sts: 28: em_un: 29: em_vi: 30: em_htg_hum: 31: em_htg_inv: 32: em_htg_other: 33: em_htgo_inv:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
------------	-------	--------------------	-------	-------------

1	72	100.0	72	AX349066	Sequence
2	68	94.4	3785	AX349063	Sequence
3	29.4	40.8	147320	AP004636	Oryza sat
4	28.6	39.7	3785	AX349063	Sequence
5	28.4	39.4	135503	AC008775	Homo sapi
6	28.4	39.4	195979	AC008519	Homo sapi
7	28.4	39.4	196599	AC037494	Homo sapi
8	27.4	38.1	168203	AC097952	Rattus no
9	27.2	37.8	8720	AC096578	Homo sapi
10	27.2	37.8	198330	AC092662	Homo sapi
11	27	37.5	102053	AC095124	Rattus no
12	27	37.5	167399	AP003236	Oryza sat
13	27	37.5	167587	AP002914	Oryza sat
14	26.8	37.2	72045	AC092214	Homo sapi
15	26.8	37.2	122717	AP001186	Homo sapi
16	26.8	37.2	156733	AP000640	Homo sapi
17	26.8	37.2	170979	AC098323	Rattus no
18	26.8	37.2	180726	AC090343	Homo sapi
19	26.8	37.2	195396	AC021809	Homo sapi
20	26.8	37.2	199801	AC093673	Homo sapi
21	26.4	36.7	200393	AC090563	Mus muscu
22	26.2	36.4	98286	AP004394	Oryza sat
23	26.2	36.4	155837	AL445495	Human DNA
24	26	36.1	55339	AC093593	Homo sapi
25	26	36.1	114745	AC027337	Homo sapi
26	26	36.1	148983	AC019019	Homo sapi
27	26	36.1	149198	AC007034	Homo sapi
28	26	36.1	160378	AC010281	Homo sapi
29	26	36.1	184210	AC011582	Homo sapi
30	26	36.1	191309	AC026711	Homo sapi
31	25.8	35.8	37840	CBRG44A09	Caenorhab
32	25.6	35.6	2571	AY051852	Drosophila
33	25.6	35.6	2573	DMU36584	Drosophila
34	25.6	35.6	3071	DR0684	Drosophila
35	25.6	35.6	21425	AC109971	Rattus no
36	25.6	35.6	75281	AL591893	Human DNA
37	25.6	35.6	143715	AL589947	Human DNA
38	25.6	35.6	155760	AC094490	Rattus no
39	25.6	35.6	160005	AC007176	Drosophila
40	25.6	35.6	161828	AC021418	Drosophila
41	25.6	35.6	168909	AC021470	Homo sapi
42	25.6	35.6	169289	AC007145	Drosophila
43	25.6	35.6	173722	AL450993	Homo sapi
44	25.6	35.6	175922	AC091242	Rattus no
45	25.6	35.6	189223	AC016773	Homo sapi

ALIGNMENTS

RESULT	1	72 bp	DNA	linear	PAT 06-FEB-2002
LOCUS	AX349066				
DEFINITION	Sequence 4 from Patent WO0202785.				
ACCESSION	AX349066				
VERSION	AX349066.1				
KEYWORDS	GI:18615101				

SOURCE

ORGANISM synthetic construct.
artificial sequence.

REFERENCE

1 (sites)
Sprunck,S., Kluth,A., Becker,D., Luetjcke,S. and Loerz,H.
Promoters of gene expression in plant caryopses
Patent: WO 0202785-A 4 10-JUN-2002;
Aventis CropScience GmbH (DE)
Location/Qualifiers

FEATURES

source

1..72
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Oligonucleotide"

BASE COUNT	15 a	20 c	17 g	20 t
ORIGIN				

Query Match 100.0%; Score 72; DB 6; Length 72;
Best Local Similarity 100.0%; Pred. No. 1.1e-16;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccttcagagacatgctgcgttccttaagacacactcttgctctatgacatgtgag 60
|||||
Db 1 cctttcagagacatgctgcgttccttaagacacactcttgctctatgacatgtgag 60

QY 61 cccaacatgtgc 72
|||||
Db 61 CCCAACAGTGC 72

RESULT 2
AX349063 3785 bp DNA linear PAT 06-FEB-2002
LOCUS AX349063
DEFINITION Sequence 1 from Patent WO0202785.
ACCESSION AX349063
VERSION AX349063.1 GI:18615098
KEYWORDS
SOURCE
ORGANISM
Triticum aestivum
bread wheat.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poidea; Triticeae; Triticum.

REFERENCE
AUTHORS Sprunck, S., Kluth, A., Becker, D., Luetticke, S. and Loeer, H.
TITLE Promoters of gene expression in plant carypsops
JOURNAL Patent: WO 0202785-A 1 10-JAN-2002;
Aventis CropScience GmbH (DE)
FEATURES
source 1..3785
Location/Qualifiers

BASE COUNT 945 a 980 c 899 g 961 t
ORIGIN
/organism="Triticum aestivum"
/db_xref="taxon:4565"

Query Match 94.4%; Score 68; DB 6; Length 3785;
Best Local Similarity 100.0%; Pred. No. 3.2e-15;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccttcagagacatgctgcgttccttaagacacactcttgctctatgacatgtgag 60
|||||
Db 1936 CCTTCAGAGACATGCTGCCTTAAGACACCTCTTGCTATGACATGTGAG 1995

QY 61 cccaacag 68
|||||
Db 1996 CCCAACAG 2003

RESULT 3
AP004636 147220 bp DNA linear HTG 17-JAN-2002
LOCUS AP004636
DEFINITION Oryza sativa chromosome 8 clone P0685B10, *** SEQUENCING IN
PROGRESS ***, in ordered pieces.
ACCESSION AP004636
VERSION AP004636.1 GI:18182016
KEYWORDS
SOURCE HTG, HTGS, PHASE2.
ORGANISM Oryza sativa (cultivar: Nipponbare) DNA, clone: P0685B10.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartioideae; Oryzae; Oryza.
1 (bases 1 to 147220)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
REFERENCE
AUTHORS Direct Submission
JOURNAL Submitted (16-JAN-2002) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program, Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail: tsasaki@nias.affrc.go.jp, URL: http://rpg.dna.affrc.go.jp/,
Tel: 81-298-38-7441, Fax: 81-298-38-7468)

COMMENT

NOTE: It currently consists of 1 config. Gaps between the configs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps between them
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source 1..147220
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="8"
/clone="P0685B10"

BASE COUNT 4225 a 3216 c 3123 g 4169 t 102 others
ORIGIN

Query Match 40.8%; Score 29.4; DB 2; Length 147220;
Best Local Similarity 76.6%; Pred. No. 1.5;
Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 17 ttgcgtgccttaagacacactcttgctctatgacatgtgagcc 63
|||||
Db 68006 TTCTCTGACTTAAACACATACCTTTGATTCATGATGTGAGCC 67960

RESULT 4
AX349063 3785 bp DNA linear PAT 06-FEB-2002
LOCUS AX349063
DEFINITION Sequence 1 from Patent WO0202785.
ACCESSION AX349063
VERSION AX349063.1 GI:18615098
KEYWORDS
SOURCE
ORGANISM
Triticum aestivum
bread wheat.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poidea; Triticeae; Triticum.

REFERENCE
AUTHORS Sprunck, S., Kluth, A., Becker, D., Luetticke, S. and Loeer, H.
TITLE Promoters of gene expression in plant carypsops
JOURNAL Patent: WO 0202785-A 1 10-JAN-2002;
Aventis CropScience GmbH (DE)
FEATURES
source 1..3785
Location/Qualifiers

BASE COUNT 945 a 980 c 899 g 961 t
ORIGIN
/organism="Triticum aestivum"
/db_xref="taxon:4565"

Query Match 39.7%; Score 28.6; DB 6; Length 3785;
Best Local Similarity 76.2%; Pred. No. 3.4;
Matches 48; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

QY 6 cagagacatgctgcgttcctt-aagacacactcttgctctatgacatgtgagcca 64
|||||
Db 2067 CGGAGCAGCCTTGCTGCTTAAGGACCTGCTTGATGATGACATGTGAGCCAG 2008

QY 65 aca 67
||
Db 2007 CCA 2005

RESULT 5
AC008775 135503 bp DNA linear PRI 15-AUG-2001
LOCUS AC008775
DEFINITION Homo sapiens chromosome 5 clone CTD-2019N2, complete sequence.
ACCESSION AC008775

VERSION	AC008775.7	GI:15187200
KEYWORDS	HTG.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
TITLE	1 (bases 1 to 135503)	
JOURNAL	DOE Joint Genome Institute and Stanford Human Genome Center.	
REFERENCE	Unpublished	
AUTHORS	2 (bases 1 to 135503)	
TITLE	DOE Joint Genome Institute.	
JOURNAL	Direct Submission	
REFERENCE	Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint	
AUTHORS	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA	
TITLE	3 (bases 1 to 135503)	
JOURNAL	DOE Joint Genome Institute and Stanford Human Genome Center.	
REFERENCE	Direct Submission	
AUTHORS	Submitted (28-FEB-2001) DOE Joint Genome Institute, 2800 Mitchell	
TITLE	Drive, Walnut Creek, CA 94598, USA	
JOURNAL	4 (bases 1 to 135503)	
REFERENCE	DOE Joint Genome Institute and Stanford Human Genome Center.	
AUTHORS	Direct Submission	
TITLE	Submitted (15-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell	
JOURNAL	Drive, Walnut Creek, CA 94598, USA	
COMMENT	On Aug 15, 2001 this sequence version replaced gi:13162501.	

FEATURES	
source	Location/Qualifiers
	1. .13503
	/Organism="Homo sapiens"
	/db_xref="taxon:9606"
	/Chromosome="5"
	/clone="CTD-2019n2"
BASE COUNT	41115 a 25039 c 25954 g 43395 t
ORIGIN	

	Query Match	Score	DB	Length	135503;
	Best Local Similarity	39.4%;	Pred.	No. 3.7'	
	Matches	44;	Conservative	0;	Mismatches 26; Indels 0; Gaps 0;
OY	2	ccttaagacagatgcttgggtgcctctaagaaccctacctttgttcatagaatgtagc	61		
Dd	3425	ctttttacagtgttgcttggstgtagacccaacacattgaattgaggctctgtgcaatttacc	3484		
OY	62	ccaacactg	71		
Dd	3485	tcaacacgtcg	3494		

RESULT	6
AC008519/c	
LOCUS	AC008519 19579 bp DNA linear PRI 23-AUG-2001
DEFINITION	Homo sapiens chromosome 5 clone CTC-460G13, complete sequence.
ACCESSION	AC008519
VERSION	AC008519.5 GI:15281177
KEYWORDS	HTG.
SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 19579)
REFERENCE	DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS	Direct Submission
TITLE	
JOURNAL	Unpublished

REFERENCE AUTHORS TITLE JOURNAL	REFERENCE AUTHORS TITLE JOURNAL	REFERENCE AUTHORS TITLE JOURNAL	REFERENCE AUTHORS TITLE JOURNAL
2 (bases 1 to 195879) DOE Joint Genome Institute. Direct Submission Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, US	3 (bases 1 to 195879) DOE Joint Genome Institute and Stanford Human Genome Center. Direct Submission Submitted (03-NOV-2000) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA	4 (bases 1 to 195879) DOE Joint Genome Institute and Stanford Human Genome Center. Direct Submission Submitted (26-NOV-2000) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA	5 (bases 1 to 195879) DOE Joint Genome Institute and Stanford Human Genome Center. Direct Submission Submitted (23-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

SHGC-10323 657536
SHGC-78846 648644.

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CFC-460G13"
BASE COUNT      64275 a 35882 c 35557 g 60265 t
ORIGIN

```

Query Match	39.4%;	Score 28.4;	DB 9;	Length 195979;
Best Local Similarity	62.9%;	Pred No. 3.7;		
Matches	44;	Conservative	0;	Mismatches 26; Indels 0; Gaps 0.

QY	2	cttcaagacgactgctcgctcctaagacaccctctgtgtctatgacatgagc	61
DB 62445	CTTTCTAGATGATGCTTGGT	TAAGACCAATACATGAAATGAGGTCTCTGTCAAATTAGC	62386
QY	62	ccacacatgag	71
DB 62385	TTCACACGTGG	62376	

[illegible]

TITLE
JOURNAL
COMMENT

Campopiano, A., Castile, A., Choepel, Y., Colangelo, M., Collins, S.,
Collimore, A., Cooke, P., Dearrellano, K., Dewar, K., Diaz, J. S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
Galagan, J., Gardyna, S., Glende, S., Coyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,
Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, D.,
Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Medcarr, J., Meneau, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tessier, S., Theodores, J., Tirrell, A., Travers, M., Triggillo, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission
Submitted (09-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 10, 2000 this sequence version replaced gi:8017214.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIMR

Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center Project name: 528_K_12

Center Clone name: 528_K_12

Sequencing vector: M13: M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 182545 bases at least Q40

Consensus quality: 189073 bases at least Q30

Consensus quality: 192263 bases at least Q20

Insert size: 208000; agarose-fp

Insert size: 194499; sum-of-ctrls

Quality coverage: 4.3 in Q20 bases; agarose-fp

Quality coverage: 4.6 in Q20 bases; sum-of-ctrls

* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 5530: contig of 5530 bp in length
* 5531 5630: gap of 100 bp
* 5631 7168: contig of 1538 bp in length
* 7169 7268: gap of 100 bp
* 7269 8574: contig of 1306 bp in length
* 8575 8674: gap of 100 bp
* 8675 9935: contig of 1261 bp in length
* 9936 10035: gap of 100 bp
* 10036 11861: contig of 1826 bp in length
* 11862 11961: gap of 100 bp
* 11962 14840: contig of 2879 bp in length
* 14841 14940: gap of 100 bp
* 14941 18392: contig of 3452 bp in length
* 18393 18492: gap of 100 bp
* 18493 22237: contig of 3745 bp in length
* 22238 22337: gap of 100 bp
* 22338 25371: contig of 3034 bp in length
* 25372 25471: gap of 100 bp
* 25472 31306: contig of 5835 bp in length
* 31307 31406: gap of 100 bp

FEATURES

source

* 31407 36289: contig of 4883 bp in length
* 36290 36389: gap of 100 bp
* 36390 42370: contig of 5981 bp in length
* 42371 42470: gap of 100 bp
* 42471 50646: contig of 8176 bp in length
* 50647 50746: gap of 100 bp
* 50747 58275: contig of 7529 bp in length
* 58276 58375: gap of 100 bp
* 58376 69446: contig of 11071 bp in length
* 69447 69546: gap of 100 bp
* 69547 82175: contig of 12629 bp in length
* 82176 82275: gap of 100 bp
* 82276 108127: contig of 25852 bp in length
* 108128 108227: gap of 100 bp
* 108228 125265: contig of 17038 bp in length
* 125266 125365: gap of 100 bp
* 125366 145214: contig of 19849 bp in length
* 145215 145314: gap of 100 bp
* 145315 161113: contig of 15799 bp in length
* 161114 161213: gap of 100 bp
* 161214 183803: contig of 22590 bp in length
* 183804 183903: gap of 100 bp
* 183904 196599: contig of 12696 bp in length.

Location/Qualifiers
1. 196599
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-528K12"
/clone_11p="RPCT-11 Human Male BAC"
1. 5330
/note="assembly-fragment
clone_end:SP6
vector_side:left"
5631. 7168
/note="assembly-fragment"
7269. 8574
/note="assembly-fragment"
8675. 9935
/note="assembly-fragment"
10036. 11861
/note="assembly-fragment"
11962. 14840
/note="assembly-fragment"
14941. 18392
/note="assembly-fragment"
18493. 22237
/note="assembly-fragment"
22338. 25371
/note="assembly-fragment"
25472. 31306
/note="assembly-fragment"
31407. 36289
/note="assembly-fragment"
36390. 42370
/note="assembly-fragment"
42471. 50646
/note="assembly-fragment"
50747. 58275
/note="assembly-fragment"
58376. 69446
/note="assembly-fragment"
69547. 82175
/note="assembly-fragment"
82276. 108127
/note="assembly-fragment"
108228. 125265
/note="assembly-fragment"
125366. 145214
/note="assembly-fragment"
145315. 161113
/note="assembly-fragment"

* arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * as soon as it is available and the accession number will
 * be preserved.

```

1 3720: contig of 3720 bp in length
3721 3820: gap of unknown length
3821 8996: contig of 5176 bp in length
8997 9096: gap of unknown length
9097 15347: contig of 6251 bp in length
15348 15447: gap of unknown length
15448 20807: contig of 5360 bp in length
20808 20907: gap of unknown length
20908 24965: contig of 4058 bp in length
24966 25065: gap of unknown length
25066 29243: contig of 4178 bp in length
29244 29343: gap of unknown length
29344 32851: contig of 3508 bp in length
32852 32951: gap of unknown length
32952 35458: contig of 2507 bp in length
35459 35558: gap of unknown length
35559 38502: contig of 2944 bp in length
38503 38602: gap of unknown length
38603 42551: contig of 3949 bp in length
42552 42651: gap of unknown length
42652 44528: contig of 1877 bp in length
44529 44628: gap of unknown length
44629 47051: contig of 2423 bp in length
47052 47151: gap of unknown length
47152 49622: contig of 2471 bp in length
49623 49722: gap of unknown length
49723 52260: contig of 2538 bp in length
52261 52360: gap of unknown length
52361 54288: contig of 1928 bp in length
54289 54388: gap of unknown length
54389 55786: contig of 1398 bp in length
55787 55886: gap of unknown length
55887 57012: contig of 1126 bp in length
57013 57112: gap of unknown length
57113 58518: contig of 1406 bp in length
58519 58618: gap of unknown length
58619 59973: contig of 1355 bp in length
59974 60073: gap of unknown length
60074 61465: contig of 1392 bp in length
61466 61565: gap of unknown length
61566 62795: contig of 1230 bp in length
62796 62895: gap of unknown length
62896 64360: contig of 1465 bp in length
64361 64460: gap of unknown length
64461 66972: contig of 2512 bp in length
66973 67072: gap of unknown length
67073 68530: contig of 1458 bp in length
68531 68630: gap of unknown length
68631 69881: contig of 1251 bp in length
69882 71210: contig of 1229 bp in length
71211 71310: gap of unknown length
71311 73418: contig of 2108 bp in length
73419 73518: gap of unknown length
73519 75998: contig of 2480 bp in length
75999 76098: gap of unknown length
76099 77417: contig of 1319 bp in length
77418 77517: gap of unknown length
77518 79011: contig of 1394 bp in length
79012 80270: contig of 1259 bp in length
80271 80370: gap of unknown length
80371 81395: contig of 1025 bp in length
81396 81495: gap of unknown length
81496 83552: contig of 1657 bp in length
83553 84749: gap of unknown length
84750 84849: gap of unknown length

```

```

* 84850 86157: contig of 1308 bp in length
* 86158 86257: gap of unknown length
* 86258 87827: contig of 1570 bp in length
* 87828 87927: gap of unknown length
* 87928 89649: contig of 1722 bp in length
* 89650 89749: gap of unknown length
* 89750 91608: contig of 1859 bp in length
* 91609 91708: gap of unknown length
* 91709 92881: contig of 1173 bp in length
* 92882 92981: gap of unknown length
* 92982 94509: contig of 1528 bp in length
* 94510 94610: gap of unknown length
* 94610 95936: contig of 1327 bp in length
* 95937 96036: gap of unknown length
* 96037 97129: contig of 1093 bp in length
* 97130 97229: gap of unknown length
* 97230 98432: contig of 1203 bp in length
* 98433 98532: gap of unknown length
* 98533 99710: contig of 1178 bp in length
* 99711 99810: gap of unknown length
* 99811 100929: contig of 1119 bp in length
* 100930 101029: gap of unknown length
* 101030 102053: contig of 1024 bp in length.

```

FEATURES
 source 1..102053
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 /clone="CH230-8N7"

BASE COUNT 27421 a 20865 c 20727 g 28460 t 4580 others

ORIGIN

Query Match 37.5%; Score 27; DB 2; Length 102053;
 Best Local Similarity 62.7%; Pred. No. 13; Indels 0; Gaps 0;
 Matches 42; Conservative 0; Mismatches 25;

```

QY 4 ttccagacagatcttcgctgccttaagacacacaccttgctgctatcacatgagccc 63
Db 20796 TTCCAGTGTAGTCTAGGTCTTATGAAATGATTCACCTACCTATGTGATTC 20737
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 64 aacagatg 70
      ||| |||
Db 20736 AACACTG 20730

```

RESULT 12
 AF003236 167399 bp DNA linear PLN 08-SEP-2001

LOCUS AF003236 Oryza sativa genomic DNA, chromosome 1, PAC clone: P0043B10.

DEFINITION AF003236

ACCESSION AF003236.3 GI:15528647

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor

Query Match	37.5%	Score 27	DB 8	Length 167399
Best Local Similarity	66.1%	Pred. No. 13		
Matches	39	Conservative	0	Mismatches 20; Indels 0; Gaps 0
QY 5	tcagagcagatcgcttcggcttaagaacccaccctctgtgtctatgacatgtgagccc 63			
Db 164257	TCAGATGAGATCCTCGACTTAACGACACAGATTGATGATCAGTATGTGGCTCC 164315			
RESULT 13				
AP002914		167587 bp	linear	PLN 29-MAR-2001
LOCUS	AP002914			
DEFINITION	Oryza sativa genomic DNA, chromosome 1, PAC clone:P0493601.			
ACCESSION	AP002914			
VERSION	AP002914.2			
KEYWORDS	GI:13486690			
SOURCE ORGANISM	Oryza sativa (cultivar:Nipponbare) DNA, clone:P0493601.			
REFERENCE	Oryza sativa			
AUTHORS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.			
TITLE	1 (sites)			
REFERENCE	Sasaki,T., Matsumoto,T. and Yamamoto,K.			
AUTHORS	Oryza sativa nipponbare(GAS) genomic DNA, chromosome 1, PAC			
TITLE	clone:P0493601			
JOURNAL	Published only in Database (2000) In press			
AUTHORS	2 (bases 1 to 167587)			
TITLE	Sasaki,T., Matsumoto,T. and Yamamoto,K.			
JOURNAL	Direct Submission			
COMMENT	Submitted (15-NOV-2000) Takuji Sasaki, National Institute of Agrobiological Resources, Rice Genome Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan (E-mail:tsasaki@affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/, Tel:81-298-38-7441, Fax:81-298-38-7468) On Mar 28, 2001 this sequence version replaced gi:11191987. Genes were predicted from the integrated results of the following: GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor (October 1998 version). The genomic sequence was searched against NCBI Nonredundant Protein database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at RGP. Protein homologs of the coding regions were searched against NCBI Nonredundant Protein database with BLASTP2.0. ESTs represent the identified cDNA sequences using BLASTN 2.0 with the corresponding DDBJ accession no. and RCP clone ID.			

A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted with a gene prediction program is classified as a 'hypothetical' protein. The orientation of the sequence is from Sp6 to T7 of the PAC clone. Detailed information on overlap and assembly quality together with annotation of this entry is available at <http://xrp.dna.affrc.go.jp/genomeseq.html>.

```

gene
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="1"
/cclone="P0493601"
/complement(join(5920, .6398,6616, .6664,6958, .7107))
/gene="P0493601.1"
/complement(join(5920, .6398,6616, .6664,6958, .7107))
/gene="P0493601.1"
/notes="contains ESTs
A0031696(K0556), C97559(C60386), C28218(C60386),
D28287(R0596)"
/codon_start=1
/product="putative glutathione S-transferase"
/protein_id="BAB39927.1"
/db_xref="GI:134866921"
/translation="MAPKVIQWVSPMARVVALBEAGAEVEYPMKSGGCHRRP
EHLANPFGIEPVLEGGDTLTLYOSALAIRIRKRPKPEFLGEGSLDESAMVDWIL
DEAHQHEAIVRPIILMHCIINKPEGRDRGVDSVEKLEKVLGYEARLGSRYTLA
GRISLDLHSHFMKRYEMATEVAGVDAYPHVKAMDEALLARPTQVKMAGMPDGG
FSGSNIP"
8240. 8389
/gene="P0493601.2"
8240. 8389
/gene="P0493601.2"
/notes="contains EST D15357(C0514)
unknown protein"
/codon_start=1
/protein_id="BAB39928.1"
/db_xref="GI:13486692"
/translation="MASPYGVYICQERHSANVFLAGCGFEEGYPVDYESALQTKAK
SYIVR"
misc_feature
complement(20654, .23458)
/gene="P0493601.3"
/notes="probably inactive due to frameshift(s) in CDS
pseudogene, 12-oxophytodienoate reductase OPR2"
pseudo
complement(20654, .23458)
/gene="P0493601.3"
complement(28447, .29904)
/gene="P0493601.4"
complement(28447, .29904)
/gene="P0493601.4"
/notes="probably inactive due to frameshift(s) in CDS
probably inactive due to no initiation/termination codons
in CDS
pseudogene, 12-oxophytodienoate reductase OPR2"
pseudo
complement(34417, .40335)
/gene="P0493601.5"
complement(34417, .40335)
/gene="P0493601.5"
/notes="probably inactive due to frameshift(s) in CDS
pseudogene, mutator-like transposase"
pseudo
49182. .50550
/gene="P0493601.6"
/notes="probably inactive due to frameshift(s) in CDS
probably inactive due to no initiation codon in CDS
probably inactive due to stop codon(s) in CDS
misc_feature

```


COMMENT

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Nov 15, 2001 this sequence version replaced gi:15625026.

----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>

Contact: sapiens@wustl.wustl.edu

----- Summary Statistics

Center project name: H_NH0811J09

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University) and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateo, M., Catanesi, J.J., and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://Daapac.med.buffalo.edu>)

VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-563K23; the clone sequenced to the right is RP11-298A10, 2000 bp overlap. Actual start of this clone is at base position 1 of RP11-811J9; actual end is at base position 140619 of RP11-298A10.

Sequence derived from PCR, base position 6122 to 6452 and 7046 to 7087.

FEATURES

Polymorphisms have been identified between AC073264 and AC092214.

SOURCE

Location/Qualifiers

1. .72045

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="7"

/map="7"

/clone="RP11-811J9"

/clone_lib="RPCI-11"

1. .437

/note="match to EST A1830201 (NID:g5450861) wj78e05.x1"

1. .333

/note="match to EST BG777585 (NID:g14047902)"

misc_feature

/note="match to EST A1066553 (NID:g3367255) ool4g02.x1"

misc_feature

/note="match to EST A1066553 (NID:g3367255) ool4g02.x1"

misc_feature

/note="match to EST BE733203 (NID:g10147195)"

misc_feature

444. .573

misc_feature

/note="similar to EST BM049883 (NID:g16779150)"

481. .482

/note="similar to Homo sapiens EST BG327560 (NID:g13133998)"

522. .525

/note="match to EST BF205674 (NID:g11099260)"

651. .678

/note="similar to Homo sapiens EST BE273823 (NID:g9148759)"

673. .773

/note="match to EST B1522595 (NID:g15347387)"

683. .773

/note="match to EST A1556892 (NID:g12899973)"

688. .773

/note="similar to Mus musculus EST BB619995 (NID:g16459189)"

691

/note="match to EST B1833448 (NID:g15944998)"

692. .773

/note="match to EST B1826228 (NID:g15937778)"

692. .742

/note="similar to Homo sapiens EST B1768766 (NID:g15760344)"

699. .773

/note="match to EST A1545268 (NID:g12877749)"

724. .773

/note="similar to Homo sapiens EST B1827429 (NID:g15938979)"

735. .773

/note="match to EST B1833448 (NID:g15944998)"

735. .773

/note="similar to Homo sapiens EST BG327560 (NID:g13133998)"

738. .773

/note="similar to Homo sapiens EST BF528670 (NID:g11616033)"

778. .837

/note="match to EST B1457310 (NID:g15247966)"

862. .1133

/note="match to EST B1457310 (NID:g15247966)"

863. .1001

/note="match to EST A1292501 (NID:g1940488) zt30b03.r1"

879. .1143

/note="match to EST BG831833 (NID:g14179420)"

921. .1143

/note="match to EST B1833448 (NID:g15944998)"

921. .1127

/note="similar to Homo sapiens EST B1827429 (NID:g15938979)"

921. .1143

/note="similar to Homo sapiens EST BG327560 (NID:g13133998)"

921. .1143

/note="match to EST B1826228 (NID:g15937778)"

923. .1143

/note="similar to Homo sapiens EST BF311877 (NID:g9148556)"

923. .1143

/note="similar to Mus musculus EST AW229324 (NID:g6558620)

924. .1143

/note="match to EST A1545268 (NID:g12877749)"

924. .1139

/note="similar to Homo sapiens EST BF528670 (NID:g11616033)"

924. .1142

/note="match to EST B1522595 (NID:g15347387)"

924. .1143

/note="match to EST A1556892 (NID:g12899973)"

924. .1143

/note="similar to Mus musculus EST BB619995 (NID:g16459189)"

939. .1143

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature


```

misc_feature /note="match to EST AU17000 (NID:910931930)"
947.1143
/note="match to EST AL54894 (NID:912884528)"
misc_feature 960.1143
/note="similar to Mus musculus EST BG663797
(NID:914214335)"
misc_feature 1001.1143
/note="similar to Mus musculus EST B1655949
(NID:915570185)"
misc_feature 1020.1143
/note="similar to Mus musculus EST B1080577
(NID:914498907)"
misc_feature 1030.1143
/note="match to EST AL048042 (NID:94728875)"

Query Match 37.2% Score 26.8; DB 9; Length 72045;
Best Local Similarity 61.4%; Pred No. 15;
Matches 43; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

OY 2 cttcaagacagatctcggtcctaagacacactctgtgtcatgacatgtgagc 61
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22320 CTTCCAGACCATTTTCATGCTCCAGAGAACTGCTGAAATGTTCTGAGATTTCCTT 22379
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 62 ccaacagctgg 71
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22380 TCACACGCG 22389

RESULT 15
AP001186 122717 bp DNA linear HTG 30-MAY-2000
LOCATION Homo sapiens chromosome 11 clone CMB9-83A3 map 11q12, WORKING DRAFT
SEQUENCE, 13 unordered pieces.
ACCESSION AP001186
VERSION HTG: HTGS.PHASE1: HTGS.DRAFT.
KEYWORDS Homo sapiens DNA, clone:CMB9-83A3.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 122717)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 122,717 genomic DNA of 11q12
TITLE Published Only in Database (2000) In press
JOURNAL 2 (bases 1 to 122717)
REFERENCE Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
AUTHORS Direct Submission
TITLE Submitted (18-FEB-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail:hattori@gs.c.riken.go.jp,
URL:http://hgp.gs.c.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)
COMMENT On May 30, 2000 this sequence version replaced gi:7023957.

----- Genome Center
Center: RIKEN Genomic Sciences Center (GSC)
Center code: RIKEN
Web site: http://hgp.gs.c.riken.go.jp/
Contact: hattori@gs.c.riken.go.jp

----- Project Information
Center project name: HumDrat11
Center clone name: CMB9-83A3
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990129
Consensus quality: 114906 bases at least Q40
Consensus quality: 118892 bases at least Q30
Consensus quality: 120476 bases at least Q20
Insert size: 121517; sum-of-contigs
Quality coverage: 6.05x in Q20 bases; sum-of-contigs

```

```

-----
NOTE: This is a 'working draft' sequence. It currently consists of
13 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved.

1 17919 contig of 17919 bp in length
18020 34939 contig of 16920 bp in length
35040 49419 contig of 14380 bp in length
49520 63383 contig of 13864 bp in length
63484 75136 contig of 11653 bp in length
75237 84053 contig of 8817 bp in length
84154 92774 contig of 7783 bp in length
92875 100657 contig of 6772 bp in length
100758 107529 contig of 5581 bp in length
107630 113210 contig of 5581 bp in length
113311 117568 contig of 4258 bp in length
117669 120668 contig of 3000 bp in length
120769 122717 contig of 1949 bp in length

Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 17919: contig of 17919 bp in length
* 17920 18019: gap of 100 bp
* 18020 34939: contig of 16920 bp in length
* 34940 35039: gap of 100 bp
* 35040 49419: contig of 14380 bp in length
* 49420 49519: gap of 100 bp
* 49520 63383: contig of 13864 bp in length
* 63384 63483: gap of 100 bp
* 63484 75136: contig of 11653 bp in length
* 75137 75236: gap of 100 bp
* 75237 84053: contig of 8817 bp in length
* 84054 84153: gap of 100 bp
* 84154 92774: contig of 8621 bp in length
* 92775 92874: gap of 100 bp
* 92875 100657: contig of 7783 bp in length
* 100658 100757: gap of 100 bp
* 100758 107529: contig of 6772 bp in length
* 107530 107629: gap of 100 bp
* 107630 113210: contig of 5581 bp in length
* 113211 113310: gap of 100 bp
* 113311 117568: contig of 4258 bp in length
* 117569 117668: gap of 100 bp
* 117669 120668: contig of 3000 bp in length
* 120669 120768: gap of 100 bp
* 120769 122717: contig of 1949 bp in length.

FEATURES
source
1. 122717
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q12"
/clone="CMB9-83A3"
1. 17919
/note="assembly-fragment clone_end:T7 vector_side:right"
18020. 34939
/misc_feature /note="assembly-fragment"
35040. 49419
/misc_feature /note="assembly-fragment"
49520. 63383
/misc_feature /note="assembly-fragment"
63484. 75136
/misc_feature /note="assembly-fragment"
75237. 84053

```

```

misc_feature      /note="assembly_fragment"
84154..92774
/note="assembly_fragment"
misc_feature      /note="assembly_fragment"
92875..100657
/note="assembly_fragment"
misc_feature      /note="assembly_fragment"
100758..107529
/note="assembly_fragment"
misc_feature      /note="assembly_fragment"
107630..113210
/note="assembly_fragment"
misc_feature      /note="assembly_fragment"
113311..117568
/note="assembly_fragment clone_end:SP6 vector_side:left"
misc_feature      /note="assembly_fragment"
117669..120668
/note="assembly_fragment"
misc_feature      /note="assembly_fragment"
120769..122717
/note="assembly_fragment"
BASE COUNT      37157 a 24437 c 24619 g 35304 t 1200 others
ORIGIN

```

```

Query Match      37.2%; Score 26.8; DB 2; Length 122717;
Best Local Similarity 68.5%; Pred. No. 15;
Matches 37; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
Qy 17 ttctgtgccttaagacaccttctgtgtctatgacatgtgagccacacagt 70
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 82525 TTtGTACCCCAAACTCTACCTTATTGTTTCAAAAGAGCCAAAAGTG 82578

```

Search completed: July 31, 2002, 14:01:44
 Job time: 17546 sec

LOCATION: 90..1472
US-08-477-347-2

Query Match 100.0%; Score 21; DB 4; Length 2224;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaaacaacaacaaca 21
|||||
DB 1847 AGAACAACAACAACA 1827

RESULT 2

US-08-476-862-1/c
; Sequence 1, Application US/08476862
; Patent No. 6262239
; GENERAL INFORMATION:

APPLICANT: WALLACH, David
APPLICANT: BIGDA, Jacek
APPLICANT: BELETSKY, Igor
APPLICANT: METT, Igor
APPLICANT: ENGELMANN, Hartmut
TITLE OF INVENTION: TNF INHIBITORS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NETMARK
STREET: 419 Seventh Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/476,862
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: IL 107267
FILING DATE: 12-OCT-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: IL 94039
FILING DATE: 06-APR-1990

PRIOR APPLICATION DATA:
APPLICATION NUMBER: IL 91229
FILING DATE: 06-AUG-1989

PRIOR APPLICATION DATA:
APPLICATION NUMBER: IL 90339
FILING DATE: 18-MAY-1989

ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: WALLACH-12A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2224 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 90..1472

US-08-476-862-1

Query Match 100.0%; Score 21; DB 4; Length 2224;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaaacaacaacaaca 21
|||||
DB 1847 AGAACAACAACAACA 1827

RESULT 3

US-09-078-294-7/c
; Sequence 7, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:

APPLICANT: Choo, Kong-Hong Andy
APPLICANT: Du Sait, Desiree
APPLICANT: Cancilla, Michael R.
TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
FILE REFERENCE: Davies Col
CURRENT APPLICATION NUMBER: US/09/078,294
CURRENT FILING DATE: 1998-05-13
NUMBER OF SEQ ID NOS: 29
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 7
LENGTH: 11811
TYPE: DNA
ORGANISM: BAC-F2 config 3
US-09-078-294-7

Query Match 100.0%; Score 21; DB 4; Length 11811;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaaacaacaacaaca 21
|||||
DB 3250 AGAACAACAACAACA 3230

RESULT 4

US-08-991-789A-191
; Sequence 191, Application US/08991789A
; Patent No. 6225054
; GENERAL INFORMATION:

APPLICANT: Frydakis, Tony N.
Reed, Steven G.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER
NUMBER OF SEQUENCES: 292
CORRESPONDENCE ADDRESS:

ADDRESSEE: Seed IP Law Group
STREET: 701 Fifth Avenue, Suite 6300
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/991,789A
FILING DATE: 11-Dec-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:

NAME: Potter, Jane E. R.
REGISTRATION NUMBER: 33,332
REFERENCE/DOCKET NUMBER: 210121.419C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

US-08-991-789A-191

INFORMATION FOR SEQ ID NO: 191;
SEQUENCE CHARACTERISTICS:
LENGTH: 369 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 191:
US-08-991-789A-191

Query Match 95.2%; Score 20; DB 4; Length 369;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaaacaacaacaaca 20
|||||
DB 148 AGAACAACAACAACA 167

RESULT 5
US-09-062-451-191
Sequence 191, Application US/09062451
Patent No. 6344550

GENERAL INFORMATION:
APPLICANT: Frudakis, Tony N.
APPLICANT: Smith, John M.
APPLICANT: Reed, Steven G.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER
NUMBER OF SEQUENCES: 297
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/062,451
FILING DATE: 04-APR-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.419C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 191:
SEQUENCE CHARACTERISTICS:
LENGTH: 369 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-062-451-191

Query Match 95.2%; Score 20; DB 4; Length 369;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaaacaacaacaaca 20
|||||
DB 148 AGAACAACAACAACA 167

RESULT 6
US-08-991-789A-169/c

Sequence 169, Application US/08991789A
Patent No. 6225054
GENERAL INFORMATION:
APPLICANT: Frudakis, Tony N.
APPLICANT: Smith, John M.
APPLICANT: Reed, Steven G.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER
NUMBER OF SEQUENCES: 292
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed IP Law Group
STREET: 701 Fifth Avenue, Suite 6300
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/991,789A
FILING DATE: 11-Dec-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Porter, Jane E. R.
REGISTRATION NUMBER: 33,332
REFERENCE/DOCKET NUMBER: 210121.419C3
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 169:
SEQUENCE CHARACTERISTICS:
LENGTH: 1265 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 169:
US-08-991-789A-169

Query Match 95.2%; Score 20; DB 4; Length 1265;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaaacaacaacaaca 20
|||||
DB 840 AGAACAACAACAACA 821

RESULT 7
US-09-062-451-169/c
Sequence 169, Application US/09062451
Patent No. 6344550

GENERAL INFORMATION:
APPLICANT: Frudakis, Tony N.
APPLICANT: Smith, John M.
APPLICANT: Reed, Steven G.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER
NUMBER OF SEQUENCES: 297
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/062,451
FILING DATE: 04-APR-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: MAKI, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.419C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 169:
SEQUENCE CHARACTERISTICS:
LENGTH: 1265 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-062-451-169

Query Match 95.2%; Score 20; DB 4; Length 1265;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaaacaacaacaaca 20
Db 840 AGAAACAACAACAACA 821

RESULT 8
US-08-991-789A-290/c
Sequence 290, Application US/08991789A
Patent No. 6225054
GENERAL INFORMATION:
APPLICANT: Fridakis, Tony N.
Smith, John M.
Reed, Steven G.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER
NUMBER OF SEQUENCES: 292
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed IP Law Group
STREET: 701 Fifth Avenue, Suite 6300
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/991,789A
FILING DATE: 11-Dec-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Potter, Jane E., R.
REGISTRATION NUMBER: 33,332
REFERENCE/DOCKET NUMBER: 210121.419C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 290:
SEQUENCE CHARACTERISTICS:
LENGTH: 1646 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 290:
US-08-991-789A-290

Query Match 95.2%; Score 20; DB 4; Length 1646;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaaacaacaacaaca 20
Db 1213 AGAAACAACAACAACA 1194

RESULT 9
US-09-062-451-290/c
Sequence 290, Application US/09062451
Patent No. 6344550
GENERAL INFORMATION:
APPLICANT: Fridakis, Tony N.
Smith, John M.
Reed, Steven G.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER
NUMBER OF SEQUENCES: 297
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/062,451
FILING DATE: 04-APR-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: MAKI, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.419C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 290:
SEQUENCE CHARACTERISTICS:
LENGTH: 1646 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-062-451-290

Query Match 95.2%; Score 20; DB 4; Length 1646;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaaacaacaacaaca 20
Db 1213 AGAAACAACAACAACA 1194

RESULT 10
US-08-967-101-86
Sequence 86, Application US/08967101
Patent No. 5840540
GENERAL INFORMATION:
APPLICANT: ST. GEORGE-HYSLOP, PETER H
ROMMENS, JOHANNA M
FRASER, PAUL E
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
TO ALZHEIMER'S DISEASE
NUMBER OF SEQUENCES: 183

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: High Street Tower - 125 High Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/967,101
FILING DATE: 10-NOV-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/592,541
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pitcher, Edmund R.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 86:
SEQUENCE CHARACTERISTICS:
LENGTH: 500 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-967-101-86

Query Match          92.4%; Score 19.4; DB 2; Length 500;
Best Local Similarity 95.2%; Pred. No. 38;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0.

QY      1 agaaacaacaacaacaacaa 21
        1 | | | | | | | | | | | | | |
Db       70 ACAACAACAACAACAACAACAA 90

RESULT 11
US-08-592-541-86
; Sequence 86, Application US/08592541
; Patent No. 5986054
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; APPLICANT: ROMMENS, JOHANNA M
; APPLICANT: FRASER, PAUL E
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
; TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 183
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: High Street Tower - 125 High Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,541
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Pitcher, Edmund R.
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 86:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 500 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-592-541-86

Query Match          92.4%; Score 19.4; DB 2; Length 500;
Best Local Similarity 95.2%; Pred No. 38;
Matches      20; Conservative    0; Mismatches     1; Indels       0; Gaps        0;

QY      1 agaacaacaacaacaacaa 21
         | ||||| ||||| ||||| ||
Db       70 ACAACAACAAACAACAA 90

RESULT 12
US-09-124-698-86
; Sequence 86, Application US/09124698
; Patent No. 6117978
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; APPLICANT: ROMMENS, JOHANNA M
; APPLICANT: FRASER, PAUL E
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
; TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 183
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HORWITZ & THIBEAULT
; STREET: High Street Tower - 125 High Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/124,698
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/592,541
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pitcher, Edmund R.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 86:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 500 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-124-698-86

Query Match          92.4%; Score 19.4; DB 3; Length 500;
Best Local Similarity 95.2%; Pred. No. 38;
Matches      20; Conservative    0; Mismatches     1; Indels       0; Gaps        0;

DB      70 ACAACAACAAACAACAA 90

```

```
RESULT 13
US-09-127-480-86
; Sequence 86, Application US/09127480
; Patent No. 6194153
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; APPLICANT: ROMMENS, JOHANNA M
; APPLICANT: FRASER, PAUL E
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
; TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 183
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: High Street Tower - 125 High Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/127,480
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,541
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pitcher, Edmund R.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 86:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 500 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-127-480-86

Query Match          92.4%; Score 19.4; DB 4; Length 500;
Best Local Similarity 95.2%; Pred. No. 38;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 agaaacaacaacaacaaca 21
DB 70 ACAACAACAACAACAACA 90

RESULT 14
US-08-496-841C-86
; Sequence 86, Application US/08496841C
; Patent No. 6210919
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; APPLICANT: ROMMENS, JOHANNA M
; APPLICANT: FRASER, PAUL E
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
; TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 175
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby, PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
```

```
ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/496,841C
; FILING DATE: 28-Jun-1995
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul F. Fehner, Ph.D.
; REGISTRATION NUMBER: 35,135
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 527-7700
; TELEFAX: (212) 753-6237
; INFORMATION FOR SEQ ID NO: 86:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 500 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 86:

US-08-496-841C-86

Query Match          92.4%; Score 19.4; DB 4; Length 500;
Best Local Similarity 95.2%; Pred. No. 38;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 agaaacaacaacaacaaca 21
DB 70 ACAACAACAACAACAACA 90

RESULT 15
US-09-328-111-120/C
; Sequence 120, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Dertl, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 120
; LENGTH: 504
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(504)
; OTHER INFORMATION: n = A,T,C or G

US-09-328-111-120
```


Thu Aug 1 08:29:59 2002

us-09-899-718a-3.rni

Page 7

	Query Match	92.4%	Score 19.4	DB 4	Length 504
	Best Local Similarity	95.2%	Pred No 38		
	Matches 20	Conservative 0	Mismatches 1	Indels 0	Gaps 0
Oy	1 agaaccaacaaacaacaa	21			
Db	402 AAAAAACAACAAACAAA	382			

Search completed: July 31, 2002, 12:20:35
Job time: 11658 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 12:11:53 ; Search time 5855.71 Seconds
(Without alignments)
48.403 Million cell updates/sec

Title: US-09-899-718a-3
Perfect score: 21
Sequence: 1 agaacaacaacaacaaca 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_esthda:*
2: em_esthum:*
3: em_estln:*
4: em_estnu:*
5: em_estoy:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
c 1	21	100.0	58	12	AZ504073	IM0344C07
c 2	21	100.0	100	10	B1046075	B1046075
c 3	21	100.0	139	10	BG988551	BG988551
c 4	21	100.0	144	10	B1043282	B1043282
c 5	21	100.0	148	9	A1345073	A1345073
c 6	21	100.0	149	10	BF922085	BF922085
c 7	21	100.0	150	9	AA639520	AA639520
c 8	21	100.0	156	12	A2493323	A2493323
c 9	21	100.0	156	12	BH105347	BH105347
c 10	21	100.0	167	9	A1253971	A1253971
c 11	21	100.0	167	9	A1733948	A1733948
c 12	21	100.0	177	9	AA719825	AA719825
c 13	21	100.0	180	9	AA457363	AA457363
c 14	21	100.0	181	9	A1733998	A1733998
c 15	21	100.0	184	9	BB072882	BB072882
c 16	21	100.0	187	9	A1251238	A1251238
c 17	21	100.0	187	9	A1792899	A1792899

c 18	21	100.0	190	12	A2656418	A2656418
c 19	21	100.0	194	9	A1549855	A1549855
c 20	21	100.0	196	9	A1792857	A1792857
c 21	21	100.0	201	9	AA839213	AA839213
c 22	21	100.0	202	9	A1465751	A1465751
c 23	21	100.0	222	9	BB150403	BB150403
c 24	21	100.0	223	9	BB536522	BB536522
c 25	21	100.0	224	9	BE071169	BE071169
c 26	21	100.0	227	9	AV317088	AV317088
c 27	21	100.0	228	9	BB183787	BB183787
c 28	21	100.0	231	9	AV229441	AV229441
c 29	21	100.0	239	10	BF747295	BF747295
c 30	21	100.0	241	12	A2221788	A2221788
c 31	21	100.0	243	9	AV281580	AV281580
c 32	21	100.0	245	9	AV262627	AV262627
c 33	21	100.0	247	9	AV230114	AV230114
c 34	21	100.0	249	10	BB852501	BB852501
c 35	21	100.0	251	9	AV336252	AV336252
c 36	21	100.0	255	12	A2414740	A2414740
c 37	21	100.0	256	9	AV264285	AV264285
c 38	21	100.0	257	10	T73967	T73967
c 39	21	100.0	260	12	A2597098	A2597098
c 40	21	100.0	262	9	BB425056	BB425056
c 41	21	100.0	264	10	BB850705	BB850705
c 42	21	100.0	267	12	A0675744	A0675744
c 43	21	100.0	268	12	A2413748	A2413748
c 44	21	100.0	270	12	A2772597	A2772597
c 45	21	100.0	275	9	BB482086	BB482086

ALIGNMENTS

RESULT 1
AZ504073/c 58 bp DNA linear GSS 05-Oct-2000
LOCUS IM0344C07F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC1M0344C07 F, DNA sequence.
ACCESSION AZ504073 GI:10685389
VERSION AZ504073.1 GI:10685389
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 58)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamli,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0344 row: C column: 07
Seq primer: CGTGTAAACGACGCCACGT
Class: plasmid ends
High quality sequence stop: 58.
Location/Qualifiers
1. 58
/organism="Mus musculus"
/strain="CS7BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0344C07"
/clone_lib="Mouse 10kb plasmid UUGC1M library"

```

/ssex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: pMD42hv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (g1/4732114/gb/AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT      11 a          9 c          10 g          28 t
ORIGIN

Query Match      100.0%; Score 21; DB 12; Length 58;
Best Local Similarity 100.0%; Pred. No. 9.6e+04;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 agaaacaaacaaacaaacaa 21
Db 23 AGAAACAAACAAACAAACAA 3

RESULT 2
LOCUS      BJ046075      100 bp      mRNA      linear      EST 06-DEC-2001
DEFINITION laevis cDNA clone XL006d02 3', mRNA sequence.
ACCESSION  BJ046075
VERSION     BJ046075.1 GI:17398266
KEYWORDS    African clawed frog.
SOURCE      Xenopus laevis
ORGANISM    Xenopus laevis
REFERENCE   1 (bases 1 to 100)
AUTHORS    Kitayama,A., Terasaka,C., Mochi,M., Ueno,N., Shin-I,T. and Kohara
Y.
TITLE      Expressed genes in X. laevis embryo
JOURNAL    Unpublished (2001)
COMMENT    Contact: Tadashi Shin-I
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshin@genes.nig.ac.jp.
            Location/Qualifiers
                1..100
                /organism="Xenopus laevis"
                /db_xref="taxon:8355"
                /clone="XL006d02"
                /clone_lib="NIBB Mochi normalized Xenopus neurula
                library"
                /tissue_type="whole embryo"
                /dev_stage="stage 15"
BASE COUNT      42 a          12 c          12 g          32 t          2 others
ORIGIN

Query Match      100.0%; Score 21; DB 10; Length 100;

```

```

Best Local Similarity 100.0%; Pred. No. 7.3e+04;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 agaaacaaacaaacaaacaa 21
Db 22 AGAAACAAACAAACAAACAA 42

RESULT 3
LOCUS      BG998551      139 bp      mRNA      linear      EST 13-JUN-2001
DEFINITION PM4-HTJ301-180401-003-g05_1 HTJ301 Homo sapiens cDNA, mRNA
sequence.
ACCESSION  BG998551
VERSION     BG998551.1 GI:14402621
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
REFERENCE   1 (bases 1 to 139)
AUTHORS    Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Joengeneel,C.V., O'Hare
            M.,J., Soares,F., Brentani,R.K., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
TITLE      Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE    20202663
COMMENT    Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM4&tl=PM4-HTJ301-
            180401-003-g05.1&tl=2001-04-18&tl=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 139.
            Location/Qualifiers
                1..139
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_lib="HTJ301"
                /dev_stage="Adult"
                /note="Organ: head_neck; Vector: puc18; Site:1; Smal;
                Site:2; SmaI; A mini-library was made by cloning products
                derived from ORESTES PCR (U.S. Letters Patent application
                No. 196,716 - Ludwig Institute for Cancer Research)
                profiles into the puc 18 vector. Reverse transcription of
                tissue mRNA and cDNA amplification were performed under
                low stringency conditions."
BASE COUNT      54 a          29 c          36 g          20 t
ORIGIN

Query Match      100.0%; Score 21; DB 10; Length 139;
Best Local Similarity 100.0%; Pred. No. 6.2e+04;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 agaaacaaacaaacaaacaa 21
Db 14 AGAAACAAACAAACAAACAA 34

RESULT 4
BASE COUNT      100.0%; Score 21; DB 10; Length 100;

```

LOCUS B1043282 144 bp mRNA linear EST 14-JUN-2001
 DEFINITION PM4-OT0201-120301-001-d03_1 OT0201 Homo sapiens cDNA, mRNA
 sequence.
 ACCESSION B1043282
 VERSION B1043282.1 GI:14450012
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 144)
 Dias Neto, E., Garcia Correa, R., Veijovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
 Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 CONTACT: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM4&t2=PM4-OT0201-
 120301-001-d03_1&t3=2001-03-12&t4=1)
 Seq primer: puc 18 forward
 High quality sequence stop: 138.
 Location/Qualifiers
 1..144
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="OT0201"
 /dev_stage="Adult"
 /note="Organ: ovary; Vector: puc18; Site_1: SmaI; Site_2:
 SmaI; A mini-library was made by cloning products derived
 from ORESTES PCR (U.S. Letters Patent application No. 196
 716 - Ludwig Institute for Cancer Research) profiles
 into the pUC 18 vector. Reverse transcription of tissue
 mRNA and cDNA amplification were performed under low
 stringency conditions."
 35 c 30 g 26 t
 BASE COUNT 53 a 35 c 30 g 26 t
 ORIGIN

Query Match 100.0%; Score 21; DB 10; Length 144;
 Best Local Similarity 100.0%; Pred. No. 6e+04;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaaacaacaacaacaaca 21
 ||||||||||||||||||||
 Db 56 AGAACAACAACAACAACA 76

RESULT 5
 LOCUS A1345073 148 bp mRNA linear EST 08-APR-1999
 DEFINITION t62b07.x1 NCI-CGAP_Brn20 Homo sapiens cDNA clone IMAGE:2058901 3',
 mRNA sequence.
 ACCESSION A1345073
 VERSION A1345073.1 GI:4082279
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 148)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncigap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgabs-r@mail.nih.gov
 Tissue Procurement: Robert Jenkins, M.D., Mark Israel, M.D., Jim
 Jacobson, Ph.D.
 CDNA Library Preparation: David B. Krizman, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/ILN at:
 www-bio.lnl.gov/bdrp/image/image.html
 Insert Length: 237 Std Error: 0.00
 Seq primer: -400p from gibco.
 Location/Qualifiers
 1..148
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="IMAGE:2058901"
 /clone_lib="NCI-CGAP_Brn20"
 /tissue_type="Oligodendroglioma"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pAMP1; mRNA made from
 Oligodendroglioma tissue, cDNA made by oligo-dt priming.
 Directionally cloned. Size selected on agarose gel,
 average insert size 500 bp. Primary library,
 non-amplified."
 26 c 30 g 29 t
 BASE COUNT 63 a 26 c 30 g 29 t
 ORIGIN

Query Match 100.0%; Score 21; DB 9; Length 148;
 Best Local Similarity 100.0%; Pred. No. 6e+04;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaaacaacaacaacaaca 21
 ||||||||||||||||||||
 Db 52 AGAACAACAACAACAACA 72

RESULT 6
 LOCUS BF922085/c 149 bp mRNA linear EST 19-JAN-2001
 DEFINITION CM2-NT0170-181100-536-e05 NT0170 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BF922085
 VERSION BF922085.1 GI:12317973
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 149)
 Dias Neto, E., Garcia Correa, R., Veijovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
 Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 CONTACT: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM2&tl2=CM2-NT0170-181100-536-e05&tl3=2000-11-18&tl4=1)
 Seq primer: puc 18 forward
 High quality sequence stop: 149.

FEATURES

source

1. 149
 Location/Qualifiers

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NT0170"

/dev_stage="Adult"

/note="Organ: nervous tumor; Vector: puc18; Site: 1; SmaI; Site: 2; SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 35 a 39 c 32 g 43 t
 ORIGIN

Query Match 100.0%; Score 21; DB 10; Length 149;
 Best Local Similarity 100.0%; Pred. No. 5.9e+04;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 agaaacaacaacaacaaa 21
 Db 62 AGAACAACAACAACAACA 42

RESULT 7
 AA639520 150 bp mRNA EST 23-OCT-1997
 LOCUS ng91a09.51 NCI_CGAP_C09 Homo sapiens cDNA clone IMAGE:1159672 3'
 DEFINITION mRNA sequence.

ACCESSION AA639520
 VERSION AA639520.1 GI:2563299
 KEYWORDS EST.

SOURCE

human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 150)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

AUTHORS

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

JOURNAL

Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: cgaaps-remail.nih.gov
 Tissue Procurement: Ilian Kirsch, M.D., Michael R. Emmert-Buck, M.D.,
 Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:

www.bio.lnl.gov/bdrip/image/image.html
 Seq primer: -40ml3 fwd. ET from Amersham.

FEATURES

source

1. 150
 Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1159672"

/clone_lib="NCI_CGAP_C09"

/tissue_type="colon tumor RER+"

/lab_host="DH10B"

/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; 1st strand cDNA was prepared from
 RER+ colon tumor, and was then primed with a Not I

oligo(dT) primer. Double-stranded cDNA was ligated to Eco

BASE COUNT 29 a 18 c 53 g 50 t
 ORIGIN

Query Match 100.0%; Score 21; DB 9; Length 150;
 Best Local Similarity 100.0%; Pred. No. 5.9e+04;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 agaaacaacaacaacaaa 21
 Db 111 AGAACAACAACAACAACA 91

RESULT 8
 AZ493323 156 bp DNA GSS 05-OCT-2000
 LOCUS 1M0327M24R Mouse 10kb plasmid UUC1M library Mus musculus genomic
 DEFINITION clone UUC1M0327M24 R. DNA sequence.

ACCESSION AZ493323.1 GI:10666890
 VERSION AZ493323
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE

1 (bases 1 to 156)
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamli, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Kelly,
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.
 and Wright, D., Weiss, R.

AUTHORS

Mouse whole genome scaffolding with paired end reads from 10kb

TITLE

Plasmid inserts

JOURNAL

Unpublished (2000)

COMMENT

Contact: Robert B. Weiss
 University of Utah
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00
 Plate: 0327 row: M column: 24
 Seq primer: CACACAGCAACACGATATGACC

Class: Plasmid ends
 High quality sequence stop: 156.
 Location/Qualifiers

FEATURES

source

1. 156
 Location/Qualifiers

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUC1M0327M24"

/clone_lib="Mouse 10kb plasmid UUC1M library"

/sex="Male"

/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: pMD42hy: Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pMD42 (g14732114|g14732072.1), a copy-number

Inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance.

BASE COUNT 28 a 41 c 29 g 58 t
ORIGIN

Query Match 100.0%; Score 21; DB 12; Length 156;
Best Local Similarity 100.0%; Pred. No. 5.8e+04;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaaacaacaacaacaaca 21
|||||
DB 113 AGAACAACAACAACAACA 93

RESULT 9

LOCUS BH105347 156 bp DNA linear GSS 19-JUL-2001
DEFINITION RPCI-24-379D17.TV RPCI-24 Mus musculus genomic clone RPCI-24-379D17
, DNA sequence.

ACCESSION BH105347
VERSION BH105347.1 GI:14934204
KEYWORDS GSS.

SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 156)
Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akiret, B., Levins, M., Tsegaye, G., Gee, K., Krol, M., Shwartsbeyn, A., Gebregeorgis, E., Russell, D., de Jong, P. and Fraser, C.M.
Mouse BAC End Sequences from library RPCI-24
Unpublished (1999)

Other GSS: RPCI-24-379D17.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

TITLE

JOURNAL

COMMENT

Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 379 row: D column: 17
Seq primer: SP6
Class: BAC ends.

FEATURES
source location/Qualifiers

1.156
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-379D17"
/clone_1ib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI; RPCI-24 Mouse BAC library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT 94 a 13 c 45 g 4 t
ORIGIN

Query Match 100.0%; Score 21; DB 12; Length 156;
Best Local Similarity 100.0%; Pred. No. 5.8e+04;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaaacaacaacaacaaca 21
|||||
DB 30 AGAACAACAACAACAACA 50

RESULT 10

LOCUS AI253971 167 bp mRNA EST 12-NOV-1998
DEFINITION qv54a05.x1 NCI_CGAP_Ov32 Homo sapiens CDNA clone IMAGE:1985360 3', mRNA sequence.

ACCESSION AI253971
VERSION AI253971.1 GI:3861496
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 167)
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps@mail.nih.gov
Tissue procurement: W. Maistron Linehan, M.D., Rodrigo F. Chuang, M.D., Michael R. Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: David B. Kitzman, Ph.D.
DNA Sequencing by: The I.M.A.G.E. Consortium, LLNL
CDNA distribution: NCI-CCGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/bbrp/image/image.html
Seq primer: 400P from Glibco.

FEATURES
source location/Qualifiers

1.167
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1985360"
/clone_1ib="NCI_CGAP_Ov32"
/sex="female"
/tissue_type="Papillary serous carcinoma"
/lab_host="DH10B"
/note="Organ: ovary; Vector: pAMP1; mRNA made from ovarian carcinoma, CDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 500 bp. Non-amplified library."
BASE COUNT 79 a 23 c 31 g 34 t
ORIGIN

Query Match 100.0%; Score 21; DB 9; Length 167;
Best Local Similarity 100.0%; Pred. No. 5.6e+04;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaaacaacaacaacaaca 21
|||||
DB 52 AGAACAACAACAACAACA 72

RESULT 11

LOCUS AI733948 167 bp mRNA EST 14-JUN-1999
DEFINITION qv54a05.x5 NCI_CGAP_Ov32 Homo sapiens CDNA clone IMAGE:1985360 3', mRNA sequence.

ACCESSION AI733948
VERSION AI733948.1 GI:5055061
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 167)
 AUTHORS NCI/NIH-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute / National Institute of Dental Research,
 Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuquai,
 M.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: David B. Krizman, Ph.D.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium, LNL
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone Distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
www.bio.lnlnl.gov/bdnp/image/image.html

This read is a RESEQUENCE of a previously sequenced human clone
 Original clone citation: National Cancer Institute, Cancer Genome
 Anatomy Project (CGAP), Tumor Gene Index
 This read has been verified (found to hit its original self in the
 correct orientation)
 Seq primer: -400P from Gibco.

FEATURES

source

1. 167

Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1985360"

/clone_lib="NCI-CGAP_Ov32"

/sex="female"

/tissue_type="papillary serous carcinoma"

/lab_host="DH10B"

/note="Organ: ovary; Vector: PAMPI; mRNA made from ovarian
 carcinoma, CDNA made by oligo-dT priming.
 Non-directionally cloned. Size-selected on agarose gel,
 average insert size 500 bp. Non-amplified library."BASE COUNT
 ORIGIN

79 a

23 c

31 g

34 t

Query Match 100.0%; Score 21; DB 9; Length 167;
 Best Local Similarity 100.0%; Pred. NO. 5.4e+04;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1 agaacacacacacacacacaa 21

DB

52 AGAACACACACACACACAA 72

RESULT 12

AA719825

LOCUS

DEFINITION

AA719825

IMAGE:414435 3'

element

AA719825.1

VERSION

AA719825

KEYWORDS

EST.

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

177 bp

mRNA

linear

EST

30-DEC-1997

1 (bases 1 to 177)

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,

Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin

J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,

White, Y., Wylie, T., Waterston, R. and Wilson, R.

WashU-NCI human EST Project

Unpublished (1997)

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800
 Fax: 314 286 1810

Email: est@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnlnl.gov) for further information.
 Seq primer: -40m13 fwd. Err from Amersham
 High quality sequence stop: 165.

FEATURES

source

1. 177

Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:414435"

/clone_lib="Soares pineal gland N3HPG"

/lab_host="DH10B (ampicillin resistant)"

/note="Organ: pineal gland; Vector: p77T3D (Pharmacia)

with a modified polylinker; Site: 1; Not I; Site 2: Eco RI;

1st strand cDNA was primed with a Not I - oligo(dT) primer

[5' TGTTACCAATCTGAAGTGGAGCGCCGCCGCTTTTCTTTTCTTTT 3']

double-stranded cDNA was size selected, ligated to Eco

RI adapters (Pharmacia), digested with Not I and cloned

into the Not I and Eco RI sites of a modified p77T3D vector

(Pharmacia). Library constructed by Bento Soares and

M. Patricia Bonaldo."

BASE COUNT
 ORIGIN

55 a

48 c

46 g

28 t

Query Match 100.0%; Score 21; DB 9; Length 177;
 Best Local Similarity 100.0%; Pred. NO. 5.4e+04;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1 agaacacacacacacacacaa 21

DB

38 AGAACACACACACACAA 58

RESULT 13

AA457363/c

LOCUS

DEFINITION

AA457363

IMAGE:838703 5'

element

AA457363.1

VERSION

AA457363

KEYWORDS

EST.

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

180 bp

mRNA

linear

EST

06-JUN-1997

1 (bases 1 to 180)

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,

Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,

Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie

T., Waterston, R. and Wilson, R.

WashU-Merck EST Project 1997

Unpublished (1997)

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.lnlnl.gov) for further information.

Seq primer: -28m13 rev. Err from Amersham

High quality sequence stop: 167.

Location/Qualifiers

1. 180

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:838703"

/clone_lib="Stratagene fetal retina 937202"

/sex="mixed"

/lab_host="SOLR (kanamycin resistant)"

/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Pooled retinal tissue. Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5' adaptor sequence: 5' GAATTCGACAGAG 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3' "

BASE COUNT 42 a 48 c 34 g 56 t

ORIGIN

Query Match 100.0%; Score 21; DB 9; Length 180;
Best Local Similarity 100.0%; Pred. No. 5.4e+04;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaacaacaacaacaaca 21
|||||
Db 171 AGAACAACAACAACAACA 151

RESULT 14

LOCUS A1733998 181 bp mRNA linear EST 13-DEC-1999
DEFINITION qv55g05.x5 NCI_CGAP_Ov32 Homo sapiens CDNA clone IMAGE:1985528 3',
MRAA sequence.

ACCESSION A1733998
VERSION A1733998.1 GI:5055111
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 181)
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuagui, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

This read is a SEQUENCE of a previously sequenced human clone
Original clone citation: National Cancer Institute, Cancer Genome
Anatomy Project (CGAP), Tumor Gene Index
This read has been verified (found to hit its original self in the
Insert length: 238 Std Error: 0.00
Seq primer: -400P from Gibco.
Location/Qualifiers
1. 181

FEATURES

source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1985528"
/clone_1lb="NCI_CGAP_Ov32"
/sex="Female"
/tissue_type="papillary serous carcinoma"
/lab_host="DH10B"
/note="Organ: ovary; Vector: PAMPI; mRNA made from ovarian carcinoma, CDNA made by oligo-dT priming.
Non-directionally cloned. Size-selected on agarose gel,
average insert size 500 bp. Non-amplified library."
BASE COUNT 78 a 32 c 34 g 37 t

Query Match 100.0%; Score 21; DB 9; Length 181;
Best Local Similarity 100.0%; Pred. No. 5.4e+04;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaacaacaacaacaaca 21
|||||
Db 75 AGAACAACAACAACAACA 95

RESULT 15

LOCUS BB072882/c 184 bp mRNA linear EST 27-JUN-2000
DEFINITION BB072882 RIKEN full-length enriched, adult male colon Mus musculus
CDNA clone 9030619B06 3' similar to U43144 Mus musculus
phospholipase C beta3 mRNA, mRNA sequence.

ACCESSION BB072882
VERSION BB072882
KEYWORDS EST.
SOURCE BB072882.1 GI:8582880

SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 184)

Authors: Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Horii, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadoya, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Satoh, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shitaki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya, T., Tsunoda, Y., Watabiki, A., Watanabe, S., Yamamura, T., Yamanaka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Konno, H., et al.)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSCC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-resgsc.riken.go.jp,
URL: <http://genome.gsc.riken.go.jp/>
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagao, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermofiltration and thermocycling for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Katsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES

source

Location/Qualifiers
1. 184
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="9030619B06"
/clone_1lb="RIKEN full-length enriched, adult male colon"
/sex="male"
/tissue_type="colon"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in

RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGATCCAGAGAGCTTTTCTTTTCTTTTCTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCGAGCTTAATTAAATTAATCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pluescript KS(+) after bulk excision from lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

BASE COUNT 43 a 37 c 45 g 59 t
ORIGIN

Query Match 100.0%; Score 21; DB 9; Length 184;
Best Local Similarity 100.0%; Pred. No. 5.3e+04;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaacacaacaacaaca 21
|||||
DB 154 AGAACCAACCAACCA 134

Search completed: July 31, 2002, 12:11:57
Job time: 13470 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 14:11:09 ; Search time 720.8 Seconds
(without alignments)
50.021 Million cell updates/sec

Title: US-09-899-718A-3

Perfect score: 21
Sequence: 1 agaacacaacaacaacaa 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: N_Geneseq_032802.*
2: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT.*
23: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT.*
24: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT.*
25: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
C 1	21	100.0	35 18 AAT93839	Phosphodiester ol1
C 2	21	100.0	140 21 AAC13041	Human secreted pro
C 3	21	100.0	384 20 AAV87203	EST clone B40. H
C 4	21	100.0	770 21 AAC44102	Arabidopsis thalia
C 5	21	100.0	1106 22 AAS46192	Human DNA encoding
C 6	21	100.0	1122 22 AAD09952	Human drug metabol
C 7	21	100.0	1409 22 ABA16382	Human nervous syst
C 8	21	100.0	1688 21 AAC79000	Human secreted pro
C 9	21	100.0	2224 16 AAO89544	p75 Tumour Necrosi

C 10	21	100.0	2339 12 AAQ10956	Encodes human 75kd
C 11	21	100.0	2339 20 AA209171	Human tumour necro
C 12	21	100.0	2339 22 AAH48860	Human TNFp-associ
C 13	21	100.0	2357 22 AAK85884	Human immune/haema
C 14	21	100.0	2358 22 AAK85885	Human immune/haema
C 15	21	100.0	2358 22 AAK85886	Human immune/haema
C 16	21	100.0	2393 12 AAQ10907	Human TNF inhibitor
C 17	21	100.0	2394 22 AAC83951	Human 40 kDa TNF i
C 18	21	100.0	2592 22 AAH15452	Human cDNA sequenc
C 19	21	100.0	2613 21 AAA49207	Human tumour necro
C 20	21	100.0	3001 21 AAH51796	Chromosome 13q31-q
C 21	21	100.0	3004 23 ABL26224	Drosophila melanog
C 22	21	100.0	4265 22 AAK53361	Human polynucleoti
C 23	21	100.0	4295 22 AAI58456	Human polynucleoti
C 24	21	100.0	4379 22 AAK53345	Human polynucleoti
C 25	21	100.0	4379 22 AAI60242	Human polynucleoti
C 26	21	100.0	4576 23 AAS84690	Human polynucleoti
C 27	21	100.0	4809 22 AAK51993	DNA encoding novel
C 28	21	100.0	5532 22 AAL04707	Human polynucleoti
C 29	21	100.0	6089 24 ABL199827	Human reproductive
C 30	21	100.0	7302 22 AAK74467	Mouse ischaemic co
C 31	21	100.0	7815 22 AAK89492	Human immune/haema
C 32	21	100.0	9328 22 AAK74469	Human immune/haema
C 33	21	100.0	10679 22 AAK89493	Human immune/haema
C 34	21	100.0	11811 20 AAV89943	Human immune/haema
C 35	21	100.0	16281 22 AAK70314	Human immune/haema
C 36	21	100.0	16285 22 AAK70315	Human immune/haema
C 37	21	100.0	16285 22 AAK70316	Human immune/haema
C 38	21	100.0	16428 22 AAK81010	Human immune/haema
C 39	21	100.0	17679 22 AAI67271	Human immune/haema
C 40	21	100.0	17679 22 AAI67272	Human immune/haema
C 41	21	100.0	19472 22 AAS26724	Human immune/haema
C 42	21	100.0	21458 22 AAL06122	Human immune/haema
C 43	21	100.0	21458 22 AAK81266	Human immune/haema
C 44	21	100.0	31834 22 AAK81009	Human immune/haema
C 45	21	100.0	31952 22 AAK89370	Human digestive sy

ALIGNMENTS

RESULT 1	
AAT93839/C	
ID AAT93839 standard; DNA; 35 BP.	
XX	
AC AAT93839;	
XX	
DT 24-FEB-1998 (first entry)	
XX	
DE Phosphodiester oligonucleotide 29 with cytotoxic activity.	
XX	
DE Phosphodiester; selective binding; cell viability; growth;	
KW tumoural cell line; cytotoxic activity; tumour cell; lymphoma;	
KW lymphoblastic tumour; ss.	
XX	
OS Synthetic.	
XX	
FT Key	Location/Qualifiers
FT modified_base 1..35	/*tag= a
FT	/note= "phosphodiester oligonucleotide"
PN	
XX	
PD W09720924-A1.	
XX	
PD 12-JUN-1997.	
XX	
PF 04-DEC-1996; 96WO-EP05388.	
XX	
PR 04-DEC-1995; 95IT-OM12539.	
XX	
PA (SAIC-) SAICOM SRL.	
XX	
PI Quadrifoglio F, Scagiantante B;	

```

XX DR WPI; 1997-319771/29.
XX PT New phospho-di:esteric oligo:nucleotide(s) - which exert a specific
XX PT and selective cytotoxic effect on tumour cells, for treating both
XX PT solid and liquid tumours
XX PS
XX PS Example 5; Page 12; 38pp; English.
CC CC Novel phosphodiesteric oligonucleotides AAT93837-41 are based on the
CC CC generic formula, in the 3'-5' or 5'-3' direction:
CC CC (Gara)'a''-(bhb)'b''-(gcgc)'c''-(gatr)'d''-(gctc)'e''-(gtrf)'f''-
CC CC (G-gtg)'g''-N', where:
CC CC N and N' = T or G, equal or different from each other;
CC CC x = 0-8, equal or different from each other;
CC CC a, b, c, d, e, f, and g = 0-10, equal or different from each other;
CC CC a', b', c', d', e', f', and g' = 0-30, equal or different from each
CC CC other;
CC CC a'', b'', c'', d'', e'', f'', and g'' = 1-16, equal or different from
CC CC each other;
CC CC The oligonucleotides (see also AAT93811-27) are believed to selectively
CC CC bind and sequester some proteins which are essential to the viability
CC CC and growth of tumoural cell lines. They have specific and selective
CC CC cytotoxic activity against tumour cells, and can be used for treating
CC CC tumours of the liquid type, in particular of lymphoblastic origin, and of
CC CC the solid type, in particular lymphomas. Oligonucleotides AAT93834-36
CC CC were created to determine the relevance of the specific features of the
CC CC above generic formula, in determining cytotoxic activity. The present
CC CC sequence is different from the generic sequence in that the sequence has
CC CC flanking fragments containing C and T bases. The results showed that
CC CC sequences having flanking fragments containing C and T bases do not exert
CC CC any significant cytotoxic activities.
XX SQ
XX SQ Sequence 35 BP; 0 A; 4 C; 5 G; 26 T; 0 other;

Query Match          100.0%; Score 21; DB 18; Length 35;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaacaacaacaacaaca 21
   |||||
Db 31 AGAACAACAACAACAACA 11

RESULT 2
AAC13041
ID AAC13041 standard; cDNA; 140 BP.
XX AC AAC13041;
XX DT 06-OCT-2000 (first entry)
XX DE Human secreted protein 5' EST, SEQ ID NO: 17116.
XX KM Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX KW gene therapy; chromosome mapping; ss.
XX OS Homo sapiens.
XX PN EP1033401-A2.
XX PD 06-SEP-2000.
XX PF 21-FEB-2000; 2000EP-0200610.
XX PR 26-FEB-1999; 99US-0122487.
XX PA (GSEST ) GENSET.
XX PI Dunas Milne Edwards J, Duclert A, Giordano J;
XX DR WPI; 2000-500381/45.

```

```

XX PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX PS
XX PS Claim 1; SEQ ID 17116; 71pp + CD-ROM; English.
CC CC The present sequence is one of a large number of 5' ESTs derived from
CC CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC CC identified within the present sequence. The 5' ESTs were prepared from
CC CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC CC derived from the 5' ends of mRNAs and even in those cases where longer
CC CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC CC They are used to obtain upstream regulatory sequences and to design
CC CC expression and secretion vectors.
XX SQ
XX SQ Sequence 140 BP; 47 A; 38 C; 38 G; 17 T; 0 other;

Query Match          100.0%; Score 21; DB 21; Length 140;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaacaacaacaacaaca 21
   |||||
Db 109 agaacaacaacaacaaca 129

RESULT 3
AAV87203
ID AAV87203 standard; cDNA; 384 BP.
XX AC AAV87203;
XX DT 27-APR-1999 (first entry)
XX DE EST clone BM40.
XX KM Expressed sequence tag; secreted protein; haematopoiesis regulator;
XX KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
XX KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
XX KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
XX OS Homo sapiens.
XX PN WO9845435-A2.
XX PD 15-OCT-1998.
XX PF 10-APR-1998; 98WO-US06954.
XX PR 10-APR-1997; 97US-0835913.
XX PA (GEMV ) GENETICS INST INC.
XX PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
XX PI Racine LA, Spaulding V, Treacy M;
XX DR WPI; 1999-070076/06.
XX PT New polynucleotides encoding human secreted proteins - derived from
XX PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
XX PT ovary, pituitary, retina and colon cDNA libraries
XX PS Claim 1; Page 487; 633pp; English.
XX CC This sequence represents an expressed sequence tag (EST), and is a

```

CC polynucleotide of the invention. The polynucleotides of the invention are
CC all secreted EST sequences isolated from a variety of human tissue
CC sources. The EST sequences and proteins encoded by them are predicted to
CC have useful biological activities which would make them suitable for
CC treating, preventing or ameliorating medical conditions in humans and
CC animals, although no supporting data is given. Suggested activities
CC include nutritional activity, immune stimulating or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity, haemostatic
CC activin/inhibin activity, chemotactic/chemokinetic activity, and inflammatory
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC therapy. The EST sequences are also stated to be useful for gene
CC
XX

SQ Sequence 384 BP; 89 A; 92 C; 80 G; 123 T; 0 other;

Query Match 100.0%; Score 21; DB 20; Length 384;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 agaacaacaacaacaaca 21
|||||

Db 216 agaacaacaacaacaaca 236

RESULT 4

AAC44102 AAC44102 standard; DNA; 770 BP.

AC AAC44102;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 41638.

KM Hybridisation assay; genetic mapping; gene expression control;

KW protein identification; signal transduction pathway;

XX metabolic pathway; promoter; termination sequence; ss.

OS Arabidopsis thaliana.

PN EPI033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132485.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.

PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151348.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0156559.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.

PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 100.0%; Score 21; DB 21; Length 770;
Best Local Similarity 100.0%; Pred. No. 1,2e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 agaacaacaacaacaaca 21
|||||

Db 11 agaacaacaacaacaaca 31

RESULT 5
AAS46192
ID AAS46192 standard; cDNA; 1106 BP.
XX
AC AAS46192;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human DNA encoding PRO polypeptide sequence #268.
XX
KW PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep; ss;
KW dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
KW blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
KW adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder;
PCR primer.
KW
XX Homo sapiens.
OS
XX WO200168848-A2.
PN
XX 20-SEP-2001.
PD
XX
PF 28-FEB-2001; 2001WO-US06520.
XX
XX 01-MAR-2000; 2000WO-US05601.
PR 02-MAR-2000; 2000WO-US05841.
PR 03-MAR-2000; 2000US-187202P.
PR 06-MAR-2000; 2000US-186968P.
PR 14-MAR-2000; 2000US-189320P.
PR 14-MAR-2000; 2000US-189328P.
PR 15-MAR-2000; 2000WO-US06884.
PR 21-MAR-2000; 2000US-190828P.
PR 21-MAR-2000; 2000US-191007P.
PR 21-MAR-2000; 2000US-191048P.
PR 21-MAR-2000; 2000US-191314P.
PR 28-MAR-2000; 2000US-192655P.
PR 29-MAR-2000; 2000US-193032P.
PR 29-MAR-2000; 2000US-193053P.
PR 30-MAR-2000; 2000WO-US08439.
PR 04-APR-2000; 2000US-194449P.
PR 04-APR-2000; 2000US-194647P.
PR 11-APR-2000; 2000US-195975P.
PR 11-APR-2000; 2000US-196000P.
PR 11-APR-2000; 2000US-196187P.
PR 11-APR-2000; 2000US-196690P.
PR 11-APR-2000; 2000US-196620P.
PR 11-APR-2000; 2000US-196820P.
PR 18-APR-2000; 2000US-198121P.
PR 18-APR-2000; 2000US-198585P.
PR 25-APR-2000; 2000US-199397P.
PR 25-APR-2000; 2000US-199550P.
PR 25-APR-2000; 2000US-199654P.
PR 25-APR-2000; 2000US-201516P.
PR 03-MAY-2000; 2000WO-US13705.
PR 17-MAY-2000; 2000WO-US14042.
PR 22-MAY-2000; 2000WO-US14042.

PR 30-MAY-2000; 2000MO-US14941.
 PR 02-JUN-2000; 2000MO-US15264.
 PR 05-JUN-2000; 2000US-209832P.
 PR 28-JUL-2000; 2000MO-US20710.
 PR 22-AUG-2000; 2000US-0644848.
 PR 24-AUG-2000; 2000MO-US23328.
 PR 08-NOV-2000; 2000MO-US30952.
 PR 01-DEC-2000; 2000MO-US32678.
 PR 20-DEC-2000; 2000MO-US34956.
 PA (GETH) GENENTECH INC.
 XX
 PI Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
 PI Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
 XX
 DR WPI; 2001-602746/68.
 DR P-PSDB; AAU29291.
 XX
 PT Novel nucleic acids encoding PRO polypeptides, used to diagnose the
 PT presence of tumours, such as prostate and breast tumours, in mammals and
 PT to screen for modulators of the compounds -
 XX
 PS Claim 2; Fig 535; 774pp; English.
 XX
 CC Sequences AAS45925-AAS46231 represent DNA molecules encoding and PCR
 CC primers for PRO polypeptides of the invention. The sequences of the
 CC invention can be used to detect the presence of a tumour in a mammal by
 CC comparing the level of expression of a PRO polypeptide in a test sample
 CC of cells from the animal and a control sample of normal cells, whereby a
 CC higher level of expression in the test sample indicates the presence of a
 CC tumour in the mammal. Mammals include dogs, cats, cattle, horses, sheep,
 CC pigs, goats and rabbits but are preferably human. The polypeptides can be
 CC used to stimulate tumour necrosis factor (TNF) alpha release from human
 CC blood, when contacted with it. A specific polypeptide can be used to
 CC stimulate the proliferation or differentiation of chondrocyte cells. The
 CC PRO proteins can be used to determine the presence of tumours and also
 CC susceptibility to tumour development, particularly adrenal, lung, colon,
 CC breast, prostate, rectal, cervical, or liver tumours. In mammalian
 CC subjects the oligonucleotide probes specific for the PRO nucleic acids
 CC can be used for genetic analysis of individuals with genetic disorders.
 CC
 SQ Sequence 1106 BP; 325 A; 226 C; 260 G; 295 T; 0 other;
 XX
 QY 1 agaacaacaacaacaaca 21
 Db 845 agaacaacaacaacaaca 865
 XX
 RESULT 6
 AAD09952
 ID AAD09952 standard; CDNA; 1122 BP.
 XX
 AC AAD09952;
 XX
 DT 12-SEP-2001 (first entry)
 DE Human drug metabolising enzyme (DME-17) cDNA.
 XX
 KW Human; drug metabolising enzyme; DME-17; immunosuppressive; gene therapy;
 KW cytosolic; autoimmune disorder; inflammatory disorder; atherosclerosis;
 KW osteoporosis; eye disorder; hepatic tumour; Addison's disease; cretinism;
 KW rheumatoid arthritis; acquired immune deficiency syndrome; AIDS; anaemia;
 KW developmental disorder; endocrine disorder; lritis; acromegaly; epilepsy;
 KW thyrotoxicosis; pancreatic disorder; diabetes mellitus; obesity; adenoma;
 KW gastrointestinal disorder; nodular hyperplasia; conjunctivitis; glaucoma;
 KW actinic keratosis; metabolic disorder; dysphagia; anorexia; carcinoma;
 XX cell proliferative disorder; ss.

OS Homo sapiens.
 XX
 FH Key
 FT CDS
 FT 123..755
 FT /*tag= a
 FT /product= "Human drug metabolising enzyme (DME-17)"
 FT 123..212
 FT /*tag= b
 FT 213..752
 FT mat_peptide
 FT /*tag= c
 FT /product= "Mature drug metabolising enzyme (DME-17)"
 XX
 XX WO200151638-A2.
 XX
 PD 19-JUL-2001.
 XX
 XX 12-JAN-2001; 2001MO-US01174.
 XX
 XX 14-JAN-2000; 2000US-0176139.
 PR 21-JAN-2000; 2000US-0177443.
 PR 28-JAN-2000; 2000US-0178574.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Yang J, Baughn MR, Burford N, Au-Young J, Lu DAM, Reddy R;
 PI Ring HZ, Hillman JL, Yue H, Azimzal Y, Tao MG, Gandhi AR;
 PI Nguyen DB, Tang YT, Lal P, Bandman O;
 XX
 DR WPI; 2001-425874/45.
 DR P-PSDB; AAE05186.
 XX
 PT Drug metabolizing enzymes and encoding polynucleotides, useful for
 PT diagnosing, treating and/or preventing autoimmune, inflammatory, cell
 PT proliferative, developmental, endocrine, eye, metabolic, and
 PT gastrointestinal disorders -
 XX
 PS Claim 5; Page 168; 133pp; English.
 XX
 CC The present sequence is human drug metabolising enzyme (DME-17) cDNA.
 CC Human DME and its nucleic acid molecule are useful for the diagnosis,
 CC treatment and prevention of disorders associated with increased or
 CC decreased expression of DME. Examples of such disorders include,
 CC autoimmune/inflammatory disorder such as acquired immune deficiency
 CC syndrome (AIDS), rheumatoid arthritis, osteoporosis; cell proliferative
 CC disorder such as actinic keratosis, atherosclerosis; developmental
 CC disorder such as epilepsy, anaemia; endocrine disorder such as
 CC acromegaly, cretinism, thyrotoxicosis; pancreatic disorder such as
 CC diabetes mellitus; eye disorder such as conjunctivitis, glaucoma, lritis;
 CC metabolic disorder such as Addison's disease, obesity; gastrointestinal
 CC disorder such as anorexia, adenomas and hepatic tumours including
 CC nodular hyperplasia, carcinomas and carcinomas.
 CC creating 'knockin' humanised animals (pigs) or transgenic animals (mice
 CC or rats) to model human disease. DME DNA is also useful for gene
 CC therapy. DME and its immunogenic fragments are useful for screening
 CC libraries of compounds in several drug screening assays.
 CC
 SQ Sequence 1122 BP; 338 A; 228 C; 261 G; 295 T; 0 other;
 XX
 QY 1 agaacaacaacaacaaca 21
 Db 845 agaacaacaacaacaaca 865
 XX
 RESULT 7
 ABA16382/c
 ID ABA16382 standard; DNA; 1409 BP.
 XX
 AC ABA16382;

XX 23-JAN-2002 (first entry)
XX Human nervous system related polynucleotide seq ID NO 8713.
DE
XX
XX Human, nootropic; neuroprotective; cytosolic; dermatological; virucide;
KW immunosuppressive; anti-inflammatory; anti-HIV; antibacterial; vulnery;
KW antiparkinsonian; antischistosomal; antitubercular; cancer;
KW antineoplastic; hepatotropic; cerebroprotective; anti-inflammatory;
KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotoxic; gene therapy; vaccine; ds.
XX Homo sapiens.
OS
XX WO200159063-A2.
PN
XX 16-AUG-2001.
PD
XX
PF 17-JAN-2001; 2001WO-US01334.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 26-JUL-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226861.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234497.
PR 25-SEP-2000; 2000US-0234498.
PR 25-SEP-2000; 2000US-0234499.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246612.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 17-NOV-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.

PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Barash SC, Ruben SM;

XX WPI; 2001-541565/60.

XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating nervous system
 PT cancers and metastases -

XX Disclosure; SEQ ID NO 8713; 1701pp + Sequence Listing; English.

XX The invention relates to novel genes (AB11004-AB21534) and proteins
 CC (AB14678-AB18001) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC disease, multiple sclerosis, rheumatoid arthritis, diabetes mellitus, Crohn's
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1409 BP; 361 A; 334 C; 324 G; 390 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 1409;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaacaacaacaacaaca 21
 ||||||||||||||||||
 DB 676 AGAAACAAACAAACAAA 656

RESULT 8

ID AAC79000 standard; DNA; 1688 BP.

XX AAC79000;

XX 14-FEB-2001 (first entry)

XX Human secreted protein gene 4 clone HCB349.

XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; anticancer;
 XX vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 XX cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 XX neurological disease; infection; human; secreted protein; ss.

XX Homo sapiens.

PN WO200058358-A1.

XX 05-OCT-2000.

XX 23-MAR-2000; 2000WO-US07725.

XX 26-MAR-1999; 99US-0126602.

XX 14-JAN-2000; 2000US-0176063.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM, Komatsu G;

XX WPI; 2000-594640/56.

XX P-PSDB; AAB44338.

XX Fourty nine nucleic acid molecules encoding human secreted proteins,
 PT useful in the prevention, treatment and diagnosis of cancer, immune
 PT disorders, cardiovascular disorders and neurological diseases -

XX Claim 1; Page 317; 367pp; English.

XX The invention relates to the isolation of genes AAC78997-C79045 encoding
 CC 49 human secreted proteins AAB44335-B44382. The genes can be used to
 CC generate fusion proteins by linking to the gene for the human
 CC immunoglobulin G Fc portion (AAC78988) for increasing the stability of
 CC the fusion protein as compared to the human protein only. The genes and
 CC proteins are useful for preventing, ameliorating or treating medical
 CC conditions, e.g. by protein or gene therapy. The genes are isolated
 CC from a range of human tissues disclosed in the specification. The
 CC nucleic acids, proteins, antibodies and (ant)agonists are useful in
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC disease, multiple sclerosis, rheumatoid arthritis, diabetes mellitus, Crohn's
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.

XX Sequence 1688 BP; 516 A; 310 C; 384 G; 478 T; 0 other;

Query Match 100.0%; Score 21; DB 21; Length 1688;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaacaacaacaacaaca 21
 ||||||||||||||||||
 DB 875 agaacaacaacaacaaca 895

RESULT 9

ID AA089544/C standard; DNA; 2224 BP.

XX AA089544;

XX 31-OCT-1995 (first entry)

XX p75 Tumour Necrosis Factor Receptor.

XX Ligand: tumour necrosis factor; nerve growth factor; TNF; NGF;

XX receptor; ss..

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 90..1475

XX /tag=a

XX /product=p75 TNF receptor.

```

FT misc_difference 1137..1139
FT /tag= b
FT /transl_except= GCA encodes Glycine.
FT misc_difference 1140..1142
FT /tag= c
FT /transl_except= CCA encodes Alanine.
FT misc_difference 1146..1148
FT /tag= d
FT /transl_except= GTG encodes Glutamic acid.
FT misc_difference 1149..1151
FT /tag= e
FT /transl_except= GAG encodes Alanine.
FT misc_difference 1152..1154
FT /tag= f
FT /transl_except= GCC encodes Arginine.
FT misc_difference 1155..1157
FT /tag= g
FT /transl_except= AGT encodes Alanine.
FT misc_difference 1158..1160
FT /tag= h
FT /transl_except= GGG encodes Serine.
FT misc_difference 1161..1163
FT /tag= i
FT /transl_except= GCC encodes Threonine.
FT misc_difference 1167..1169
FT /tag= j
FT /transl_except= GAG encodes Serine.
FT misc_difference 1170..1172
FT /tag= k
FT /transl_except= GCC encodes Serine.
FT misc_difference 1173..1175
FT /tag= l
FT /transl_except= CGG encodes Aspartic acid.
FT misc_difference 1176..1178
FT /tag= m
FT /transl_except= GCC encodes Serine.
FT misc_difference 1182..1184
FT /tag= n
FT /transl_except= ACC encodes Proline.
FT misc_difference 1188..1190
FT /tag= o
FT /transl_except= AGC encodes Glycine.
FT misc_difference 1191..1193
FT /tag= p
FT /transl_except= TCA encodes Histidine.
FT misc_difference 1194..1196
FT /tag= q
FT /transl_except= GAT encodes Glycine.
FT misc_difference 1197..1199
FT /tag= r
FT /transl_except= TCT encodes Threonine.
FT misc_difference 2000..2002
FT /tag= s
FT /transl_except= TCC encodes Glutamine.
FT misc_difference 2003..2005
FT /tag= t
FT /transl_except= CCT encodes Alanine.
FT misc_difference 2006..2008
FT /tag= u
FT /transl_except= GGT encodes Proline.
FT misc_difference 2012..2014
FT /tag= v
FT /transl_except= CAT encodes Valine.
FT misc_difference 2015..2016
FT /tag= w
FT /transl_except= GGG encodes Glutamic acid.
FT misc_difference 2017..2018
FT /tag= x
FT /transl_except= ACC encodes Alanine.
FT misc_difference 2019..2021
FT /tag= y
FT /transl_except= CAG encodes Serine.
XX

```

```

PN EP648783-A.
XX
XX PD 19-APR-1995.
XX 11-OCT-1994; 94EP-0116015.
XX PR 12-OCT-1993; 93II-0107267.
XX PA (YEDA ) YEDA RES & DEV CO LTD.
XX (WALL/) WALLACH D.
XX PI Beletsky I, Bigda J, Mett I, Wallach D;
XX WPI: 1995-148673/20.
XX DR P-PSDB; AAR72504.
XX PT Tumour necrosis factor (TNF) receptor ligand - used to increase
XX inhibitory effect of a soluble TNF receptor
XX PS Disclosure; Figure 2; 18pp; English.
XX CC A ligand to a member of the tumour necrosis factor (TNF)/nerve
XX growth factor (NGF) receptor family which binds either to the region
XX of the 4th-Cys rich domain of the receptor, or to the region between
XX it and the cell membrane may be used in the production of a
XX pharmaceutical composition for increasing the inhibitory effect of a
XX soluble receptor of the TNF/NGF receptor family. This sequence
XX encodes the p75 TNF receptor. N in the sequence represents an
XX unidentified nucleotide (poor reproduction in specification).
SQ Sequence 2224 BP; 432 A; 697 C; 688 G; 400 T; 7 other;

```

```

QY 1 agaaacaacaacaacaaca 21
DB 1847 AGAACAACAACAACAACA 1827

RESULT 10
AAQ10956/C
ID AAQ10956 standard; DNA; 2339 BP.
XX
AC AAQ10956;
XX
DT 24-MAY-1991 (first entry)
XX
DE Encodes human 75kd TNF-binding protein.
XX
XX Tumour Necrosis Factor; binding proteins; septic shock;
KW autoimmune glomerulonephritis; lymphokine; cytokine.
XX
FH Key 1..1179
FT CDS /tag= a
FT /product= 75kd TNF-BP
FT
XX EP417563-A.
XX
XX 20-MAR-1991.
XX
XX 31-AUG-1990; 90EP-0116707.
XX
XX 20-APR-1990; 90CH-0001347.
XX 12-SEP-1989; 89CH-0003319.
XX 08-MAR-1990; 90CH-0000746.
XX
XX (HOFF ) HOFFMANN-LA ROCHE AG.
XX
XX Brockhaus M, Dembic Z, Gentz R, Lesslauer W, Lotscher H;

```

PI Schlaeger EJ;
 XX WPI; 1991-081851/12.
 DR P-PSDB; AAR11605.
 XX
 PT Insoluble tumor necrosis factor binding proteins - and DNA
 PT encoding them, useful in pharmaceutical prods. and for antibody
 PT prodn.
 PS
 XX Claim 4; Fig 1; 26pp; German.
 XX
 CC Partial amino acid sequences were determined for the 55 and 75KD
 CC TNF-BPs (see AAR11072-R11081) and oligonucleotide primers were
 CC synthesised based on these partial sequences. The primers were
 CC to produce a cDNA fragment for use as a probe to screen a human
 CC placental cDNA bank constructed in lambda gt11. Positive clones were
 CC identified and sequenced. Repeated sequencing showed a discrepancy
 CC at position 7 such that the third codon encodes either Thr or Ser.
 CC DNA constructs comprising the TNF-BP coding sequence may also
 CC contain a fragment encoding a human Ig domain. Recombinant
 CC constructs are used to transform cells to confer improved TNF-
 CC binding properties.
 CC See also AAQ10955.
 CC
 SQ Sequence 2339 BP; 494 A; 720 C; 685 G; 439 T; 1 other;

Query Match 100.0%; Score 21; DB 12; Length 2339;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 agaaacaacaacaacaaca 21
 ||||||||||||||||||||
 Db 1550 AGAAACAAACAAACAAACAA 1530

RESULT 11
 AAZ09171/C
 ID AAZ09171 standard; cDNA; 2339 BP.
 XX
 AC AAZ09171;
 DT 18-OCF-1999 (first entry)
 XX
 DE Human tumour necrosis factor binding protein cDNA fragment.
 XX
 KW Tumour necrosis factor binding protein; TNF; insoluble protein; agonist;
 KW anti-inflammatory; antimalarial; treatment; septic shock; inflammation;
 KW autoimmune glomerulonephritis; cerebral malaria; immune response;
 KW antagonist; diagnosis; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT 1..1179
 FT CDS /*tag= a
 FT /product= "TNF binding protein"
 FT /note= "Partial sequence, no start codon given"
 XX
 PN EP939121-A2.
 PD 01-SEP-1999.
 XX
 PF 31-AUG-1990; 90EP-0116707.
 PR 20-APR-1990; 90CH-0001347.
 PR 12-SEP-1989; 89CH-0003319.
 PR 08-MAR-1990; 90CH-0000746.
 XX
 PA (HOFF) HOFFMANN LA ROCHE & CO AG F.
 XX Brockhaus M, Dembic Z, Gentz R, Lesslauer W, Loetscher H;
 PI Schlaeger E;

XX
 DR WPI; 1999-480840/41.
 DR P-PSDB; AAV30935.
 XX
 PT New insoluble proteins, and fragments, that bind to tumor necrosis
 PT factor, used to treat e.g. septic shock or cerebral malaria
 XX
 PS Claim 4a; Fig 4; 25pp; German.
 XX
 CC This invention describes novel homogeneous insoluble proteins (I),
 CC their (in)soluble fragments (Ia) and their salts that can bind tumor
 CC necrosis factor (TNF). The products of the invention have
 CC anti-inflammatory and antimalarial activity. (I) and (Ia) are used (i)
 CC to treat diseases in which TNF is involved (e.g. septic shock, autoimmune
 CC glomerulonephritis, cerebral malaria, immune responses and inflammation),
 CC (ii) to purify TNF, (iii) to identify TNF (ant)agonists and (iv) for
 CC diagnostic determination of TNF in body fluids. Antibodies raised against
 CC (I) are used for affinity purification of (I). This sequence encodes
 CC a tumor necrosis factor binding protein fragment described in the method
 CC of the invention.
 CC
 SQ Sequence 2339 BP; 494 A; 720 C; 685 G; 440 T; 0 other;

Query Match 100.0%; Score 21; DB 20; Length 2339;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 agaaacaacaacaacaaca 21
 ||||||||||||||||||||
 Db 1550 AGAAACAAACAAACAAACAA 1530

RESULT 12
 AAH48860/C
 ID AAH48860 standard; DNA; 2339 BP.
 XX
 AC AAH48860;
 DT 12-NOV-2001 (first entry)
 XX
 DE Human TNFBP-associated DNA #2.
 XX
 KW TNF; tumor necrosis factor binding protein; TNFBP; treatment;
 KW insoluble protein; anti-inflammatory; immunosuppressive; antibacterial;
 KW antiprotozoal; treatment; meningococcal sepsis; cerebral malaria;
 KW autoimmune glomerulonephritis; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT 1..1179
 FT CDS /*tag= a
 FT /product= "TNFBP-associated protein"
 XX
 PN EP1132471-A2.
 PD 12-SEP-2001.
 XX
 PF 31-AUG-1990; 2001EP-0108117.
 PR 12-SEP-1989; 89CH-0003319.
 PR 08-MAR-1990; 90CH-0000746.
 PR 20-APR-1990; 90CH-0001347.
 PR 31-AUG-1990; 90EP-0116707.
 PR 31-AUG-1990; 99EP-0100703.
 XX
 PA (HOFF) HOFFMANN LA ROCHE & CO AG F.
 XX Brockhaus M, Dembic Z, Gentz R, Lesslauer W, Loetscher H;
 PI Schlaeger E;
 XX WPI; 2001-559312/63.

DR P-PSDB; AAB86818.
XX
PR New homogeneous, insoluble proteins that bind tumor necrosis factor
PR (TNF), useful for treating TNF-mediated disorders, e.g. inflammation
XX
PS Claim 4a; Fig 4; 26pp; German.
XX
CC This invention describes novel insoluble proteins (I), also their
CC (in)soluble fragments and pharmaceutically acceptable salts, able to bind
CC tumor necrosis factor (TNF) and in homogeneous form. The products of the
CC invention have antiinflammatory, immunosuppressive, antibacterial,
CC antiprotocoll activity. (I), and related recombinant proteins, are used
CC to treat diseases mediated by TNF, e.g. shock in cases of meningococcal
CC sepsis; development of autoimmune glomerulonephritis and cerebral
CC malaria. Also (I), or antibodies specific for them, are used for
CC diagnostic determination of TNF in body fluids, for affinity purification
CC of TNF and for identifying (ant)agonists of TNF. This sequence encodes a
CC human TNF binding protein described in the method of the invention.
CC
XX
SQ Sequence 2339 BP; 494 A; 720 C; 685 G; 440 T; 0 other:

Query Match 100.0%; Score 21; DB 22; Length 2339;
Best Local Similarity 100.0%; Pred. No. 1; le+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 agaacaacaacaacaaca 21
|||||
DB 1550 AGAACAACAACAACAACA 1530

RESULT 13
AAK85884/C
ID AAK85884 standard; DNA; 2357 BP.
XX
AC AAK85884;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:40696.
XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226881.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234224.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0238935.
PR 13-OCT-2000; 2000US-0238937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.

PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246529.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249254.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250191.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251889.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis.
XX
XX Disclosure; SEQ ID NO 40696; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
XX amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patients own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting
XX the nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX cancers and cancer metastases of haematopoietic-related diseases, especially
XX to AAK87694 represent human immune/haematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAM82169
XX represent sequences used in the exemplification of the present invention.

Sequence 2357 BP; 677 A; 465 C; 347 G; 868 T; 0 other;
Query Match 100.0%; Score 21; DB 22; Length 2357;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 agaacaacaacaacaaca 21
DB 427 AGAACAACAACAACAACA 407
RESULT 14
AAK85885/C
ID AAK85885 standard; DNA; 2358 BP.
XX
XX AAK85885;
XX
XX 07-NOV-2001 (first entry)
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:40697.
XX
XX DE Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX OS Homo sapiens.
XX
XX PN WO200157182-A2.
XX
XX PD 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225216.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0226868.
XX 22-AUG-2000; 2000US-0227182.
XX 23-AUG-2000; 2000US-0227009.
XX 30-AUG-2000; 2000US-0228924.
XX 01-SEP-2000; 2000US-0229287.
XX 01-SEP-2000; 2000US-0229343.
XX 01-SEP-2000; 2000US-0229344.
XX 01-SEP-2000; 2000US-0229345.

	PR	17-NOV-2000;	2000US-0249216.	
	PR	17-NOV-2000;	2000US-0249217.	
	PR	17-NOV-2000;	2000US-0249218.	
	PR	17-NOV-2000;	2000US-0249244.	
	PR	17-NOV-2000;	2000US-0249245.	
	PR	17-NOV-2000;	2000US-0249264.	
	PR	17-NOV-2000;	2000US-0249265.	
	PR	17-NOV-2000;	2000US-0249297.	
	PR	17-NOV-2000;	2000US-0249300.	
	PR	17-NOV-2000;	2000US-0250160.	
	PR	01-DEC-2000;	2000US-0250391.	
	PR	05-DEC-2000;	2000US-0251030.	
	PR	05-DEC-2000;	2000US-0251988.	
	PR	05-DEC-2000;	2000US-0251679.	
	PR	06-DEC-2000;	2000US-0251856.	
	PR	08-DEC-2000;	2000US-0251868.	
	PR	08-DEC-2000;	2000US-0251869.	
	PR	08-DEC-2000;	2000US-0251989.	
	PR	08-DEC-2000;	2000US-0251990.	
	PR	11-DEC-2000;	2000US-0254097.	
	PR	05-JAN-2001;	2001US-0259678.	
XX	(HUMA -) HUMAN GENOME SCI INC.			
PA	Rosen CA,	Barash SC,	Ruben SW:	
XX				
DR	WPI; 2001-483426/52.			
PT	Nucleic acids encoding human immune/hematopoietic antigen polypeptides.			
PM	useful for preventing, diagnosing and/or treating cancers and			
PT	metastasis -			
PS	Disclosure; SEQ ID NO 40697; 3071pp + Sequence Listing; English.			
CC	AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)			
CC	amino acid sequences given in AAM82170 to AAM91921. (I) have cytosolic			
CC	'activity' and can be used in gene therapy and vaccine production. (I)			
CC	proteins and polynucleotides may be used in the prevention, diagnosis and			
CC	treatment of diseases associated with inappropriate (I) expression. For			
CC	example, they may be used to treat disorders associated with decreased			
CC	expression by rectifying mutations or deletions in a patient's genome			
CC	that affect the activity of (I) by expressing inactive proteins or to			
CC	suppress the patients own production of (I). Additionally, (I)			
CC	polynucleotides may be used to produce the secreted (I), by inserting			
CC	the nucleic acids into a host cell and culturing the cell to express the			
CC	protein. (I) proteins and polynucleotides may be used to prevent,			
CC	diagnose and treat immune/hematopoeitic-related diseases, especially			
CC	cancers and cancer metastases of hematopoietic-derived cells. AAK64703			
CC	to AAK87694 represent human immune/hematopoietic antigen genomic			
CC	sequences from the present invention. AAK54947 to AAK54950 and AAM82169			
CC	represent sequences used in the exemplification of the present invention.			
XX				
SS	Sequence 2358 BP; 679 A; 467 C; 345 G; 867 T; 0 other;			
	Query Match	100.0%; Score 21; DB 22; Length 2358;		
	Best Local Similarity	100.0%; Pred. No. 1.le+02;		
	Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0			
DG	1 agaaacaacaaaacaaaaa 21 			
ID	427 AGAACAACCAACCAACAAA 407			
RESULT 15				
AAR8586/C				
ID	AAR8586 standard; DNA; 2358 BP.			
AC	AAR8586;			
DT	07-NOV-2001 (first entry)			


```

PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 03-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis -
XX
XX PS Disclosure; SEQ ID NO 40698; 3071pp + Sequence Listing; English.
XX
XX CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
XX amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patients own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting
XX the nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/hematopoietic-related diseases, especially
XX cancers and cancer metastases of hematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/hematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAM82169
XX represent sequences used in the exemplification of the present invention.
XX
XX SQ Sequence 2358 BP; 679 A; 464 C; 347 G; 868 T; 0 other;

```

```

Query Match 100.0%; Score 21; DB 22; Length 2358;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 agaaacaacaacaacaaca 21
   ||||||||||||||||
DB 427 AGAACAACAACAACAACA 407

```

Search completed: July 31, 2002, 14:11:11
 Job time: 17368 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 14:00:49 ; Search time 6034.22 Seconds
(without alignments)
72.828 Million cell updates/sec

Title: US-09-899-718A-3

Perfect score: 21

Sequence: 1 aagaacaacaacaacaaca 21

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 segs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba: *
2: gb_hcg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vl: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htgo_inv: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Match length	ID	Description
------------	-------	--------------	----	-------------

1	21	100.0	21	6	AX349065	AX349065 Sequence
2	21	100.0	126	10	RNU76838	U76838 Rattus norv
3	21	100.0	110	10	RNU76839	U76839 Rattus norv
4	21	100.0	224	8	AY023679	AY023679 Oryza sat
5	21	100.0	239	11	G35045	G35045 STS NR03-14
6	21	100.0	281	11	AU026129	AU026129 Rattus no
7	21	100.0	309	9	HUMRPPLM	L20486 Human micro
8	21	100.0	381	11	HUMUT925	L16357 Human chrom
9	21	100.0	407	11	HUMUT1400	L17810 Human STS U
10	21	100.0	497	9	HUMDS320T	L12265 Homo sapien
11	21	100.0	497	11	HUMUT5307	L18590 Human chrom
12	21	100.0	503	11	G24380	G24380 human STS W
13	21	100.0	552	11	G42246	G42246 58P187 Hum
14	21	100.0	561	11	HUMUT6478	L30582 Human STS U
15	21	100.0	617	5	AF254660	AF254660 Tetrao ur
16	21	100.0	729	9	HS297552	AJ287552 Homo sapi
17	21	100.0	766	8	AY054247	AY054247 Arabidops
18	21	100.0	772	10	MMU57395	U57395 Mus musculu
19	21	100.0	804	9	HSNMF4GD01	AF133493 Homo sapi
20	21	100.0	870	11	G15915	G15915 human STS C
21	21	100.0	994	10	MMU91A5	AF040950 Mus muscu
22	21	100.0	1122	6	AX195188	AX195188 Sequence
23	21	100.0	1197	10	AF19900382	AF199004 Mus muscu
24	21	100.0	1213	10	RATGSTPDS	M14364 Rat glutath
25	21	100.0	1687	5	XL038844	U38844 Xenopus lae
26	21	100.0	1903	5	GGU31223	U31223 Gallus gall
27	21	100.0	1970	17	RATR2RB	L14610 Rat transcr
28	21	100.0	2084	5	CCJUN	X15547 Quail mRNA
29	21	100.0	2224	6	AR152033	AR152033 Sequence
30	21	100.0	2253	6	A78517	A78517 Sequence 1
31	21	100.0	2339	6	A26415	A26415 CDNA fragme
32	21	100.0	2390	10	MMMH07G1	X03210 Mouse MHC Q
33	21	100.0	2394	9	HUMTNPRII	M55994 Human tumor
34	21	100.0	2592	9	AK021890	AK021890 Homo sapi
35	21	100.0	2613	9	HSTNFK2S10	U52165 Human tumor
36	21	100.0	2831	8	PBL276965	AJ276965 Phycomyce
37	21	100.0	3100	8	PBL278287	AJ278287 Phycomyce
38	21	100.0	3380	11	G26865	G26865 human STS S
39	21	100.0	3492	21	S63368	S63368 Homo sapien
40	21	100.0	3588	10	AB044626	AB044626 Mus muscu
41	21	100.0	3683	6	AX333705	AX333705 Sequence
42	21	100.0	3683	6	AX348016	AX348016 Sequence
43	21	100.0	3683	6	AX348018	AX348018 Sequence
44	21	100.0	3683	6	AX348020	AX348020 Sequence
45	21	100.0	3683	9	HUMNFR	M32315 Human tumor

ALIGNMENTS

RESULT	1	AX349065	Sequence 3 from Patent WO0202785.	21 bp	DNA	linear	PAT 06-FEB-2002
LOCUS	AX349065						
DEFINITION	AX349065						
ACCESSION	AX349065.1	GI:18615100					
VERSION							
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
AUTHORS							
TITLE							
JOURNAL							
FEATURES							
Source							

synthetic construct.
artificial sequence.
1 (sites)
Sprunck,S., Kluth,A., Becker,D., Luetticke,S. and Loez,H.
Promoters of gene expression in plant carypsops
Patent: WO 0202785-A 3 10-JAN-2002;
Aventis CropScience GmbH (DE)
Location/Qualifiers
1..21
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Oligonucleotide"

BASE COUNT	16 a	4 c	1 g	0 t
ORIGIN				

Query Match 100.0%; Score 21; DB 6; Length 21;
 Best Local Similarity 100.0%; Pred. No. 3e+02;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 agaaacaacaacaacaaca 21
 ||||||||||||||||||
 Db 1 AGAACAACAACAACAACA 21

RESULT 2

RNU76839 116 bp DNA linear ROD 27-SEP-1997
 LOCUS Rattus norvegicus cyclin A gene, intronic microsatellite repeat
 DEFINITION region.

ACCESSION U76838
 VERSION U76838.1 GI:2443415

SOURCE

Norway rat.
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE 1 (bases 1 to 116)
 BROCKMANN, B., SCHNIEDERS, F., KUNZE, B., KRAWCZAK, M. and SCHMIDTKE, J.
 TITLE DNA sequence divergence in a cyclin A gene of Apodemus sylvaticus
 and Apodemus flavicollis

JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 116)
 BROCKMANN, B., SCHNIEDERS, F., KUNZE, B., KRAWCZAK, M. and SCHMIDTKE, J.
 AUTHORS Direct Submission
 TITLE Submitted (01-NOV-1996) Max-Planck-Group, MDC for Molecular
 JOURNAL Medicine, Robert-Rossle-Str. 10, Berlin-Buch 13125, Germany

FEATURES

source 1..116
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 intron <1..>116
 /note="cyclin A; Allele: 1"
 repeat_region 80..116
 /note="intronic"
 /rpt_family="microsatellite"
 BASE COUNT 58 a 17 c 19 g 22 t
 ORIGIN

Query Match 100.0%; Score 21; DB 10; Length 116;
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 agaaacaacaacaacaaca 21
 ||||||||||||||||||
 Db 94 AGAACAACAACAACAACA 114

RESULT 3

RNU76839 120 bp DNA linear ROD 27-SEP-1997
 LOCUS Rattus norvegicus cyclin A gene, intronic microsatellite repeat
 DEFINITION region.

ACCESSION U76839
 VERSION U76839.1 GI:2443416

KEYWORDS

SOURCE

Norway rat.
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE 1 (bases 1 to 120)
 BROCKMANN, B., SCHNIEDERS, F., KUNZE, B., KRAWCZAK, M. and SCHMIDTKE, J.
 AUTHORS DNA sequence divergence in a cyclin A gene of Apodemus sylvaticus
 TITLE and Apodemus flavicollis
 JOURNAL Unpublished

REFERENCE 2 (bases 1 to 120)
 BROCKMANN, B., SCHNIEDERS, F., KUNZE, B., KRAWCZAK, M. and SCHMIDTKE, J.
 TITLE Direct Submission
 JOURNAL Submitted (01-NOV-1996) Max-Planck-Group, MDC for Molecular
 Medicine, Robert-Rossle-Str. 10, Berlin-Buch 13125, Germany

FEATURES
 source 1..120
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"

intron

<1..>120
 /note="cyclin A; Allele: 2"

repeat_region 80..120
 /note="intronic"
 /rpt_family="microsatellite"

BASE COUNT 61 a 17 c 20 g 22 t
 ORIGIN

Query Match 100.0%; Score 21; DB 10; Length 120;
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 agaaacaacaacaacaaca 21
 ||||||||||||||||||
 Db 98 AGAACAACAACAACAACA 118

RESULT 4

AY023679 224 bp DNA linear. PLN 07-FEB-2001
 LOCUS Oryza sativa microsatellite MR6004 containing (AAAC)X6, closest to
 DEFINITION marker R1479, genomic sequence.

ACCESSION AY023679
 VERSION AY023679.1 GI:12706895

KEYWORDS

SOURCE

Oryza sativa.
 Oryza sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzoae; Oryza.

REFERENCE 1 (bases 1 to 224)
 TAO, N., BARBAZUK, W.B., LIU, J., WU, K. and BARRY, G.F.
 TITLE Simple sequence repeats from Monsanto rice genomic sequences
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 224)
 TAO, N., BARBAZUK, W.B., LIU, J., WU, K. and BARRY, G.F.
 TITLE Direct Submission

COMMENT Submitted (10-JAN-2001) Genomics, Monsanto, 800 North Lindbergh
 Blvd., Creve Coeur, MO 63167, USA
 Derived from rice genomic sequences generated from the Monsanto
 Rice Genome Sequencing project. Please see
 http://www.rice-research.org for more information. The sequence
 data were produced primarily in the laboratories of Dr. Leroy Hood
 at the University of Washington in Seattle.

FEATURES

source 1..224
 /organism="Oryza sativa"
 /db_xref="taxon:4530"
 repeat_region 1..224
 /note="microsatellite MR6004"
 /rpt_type=tandem
 /rpt_unit=aaac

BASE COUNT 83 a 53 c 45 g 43 t
 ORIGIN

Query Match 100.0%; Score 21; DB 8; Length 224;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 agaaacaacaacaacaaca 21
 ||||||||||||||||||
 Db 99 AGAACAACAACAACAACA 119

RESULT 5
G35045
LOCUS STS NRU3-14s clone NRU3-14, sequence tagged site.
DEFINITION
ACCESSION G35045
VERSION G35045.1 GI:2358168
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE human vector-pbluescript host-XL1 Blue Alu PCR products of chromosome 3 radiation hybrids cloned into pbs.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 239)
AUTHORS Naylor,S.L., Garcia,D.K., Kerbacker,K., O'Connell,P. and Stanton,V.
TITLE Human Chromosome 3 STS
JOURNAL Unpublished
COMMENT
GDB_DSEG: D3S4038
Contact: Susan L. Naylor
San Antonio Genome Center
Univ. Of TX Health Science Center, San Antonio
7703 Floyd Curl Dr. San Antonio, TX 78385, USA
Tel: (210)567-3842
Fax: (210)567-6781
Email: naylor@uthscsa.edu
Primer A: AACACTATTCCTCCAC
Primer B: CCGCATATAACTTCAGTCC
STS size: 132
PCR Profile:
Presoak: 95 degrees C for 5 minutes
Denaturation: 95 degrees C for 1 minute
Annealing: 56 degrees C for 1 minute
Extension: 72 degrees C for 1 minute
PCR cycles: 30
Final extension: 72 degrees C for 10 minutes
Thermal cycler: Perkin Elmer 48
Protocol:
Template: 200ng
Primer: 10ng each
dNTPs: 4mM of each dNTP/20ul rxn
Taq polymerase: 1 unit /20ul rxn
Total Vol: 20ul
Buffer:
MgCl2: 1.5mM
KCl:
Tris-HCl:
pH:
FEATURES
source Location/Qualifiers
1..239
/organism="Homo sapiens"
/db_xref="taxon:9606"
STS
83..214
primer_bind 83..100
primer_bind complement(195..214)
BASE COUNT 80 a 56 c 39 g 64 t
ORIGIN
Query Match 100.0%; Score 21; DB 11; Length 239;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 11 AGAACAACAACAACAACA 31

RESULT 6
AU026129/c

LOCUS AU026129 281 bp DNA linear STS 02-MAR-1999
DEFINITION Rattus norvegicus, OTSUKA clone, OT58.09/882h07, microsatellite
sequence, sequence tagged site.
ACCESSION AU026129
VERSION AU026129.1 GI:4516052
KEYWORDS STS.
SOURCE Rattus norvegicus DNA, clone:OT58.09/882h07.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (sites)
AUTHORS Watanabe,T.K., Hishigaki,H., Kanemoto,N., Miyakita-Mizoguchi,A., Oga,K., Okuno,S., Ono,T., Tsuji,A., Hayashi,H., Adachi,M., Yamasaki,Y., Iriye,Y., Takahashi,E., Takagi,T., Nakamura,Y. and Tanigami,A.
TITLE The large-scale mapping of rat microsatellite markers
JOURNAL Unpublished (1998)
REFERENCE 2 (bases 1 to 281)
AUTHORS Watanabe,T.K.
TITLE Direct Submission
JOURNAL Submitted (24-JUL-1998) to the DDBJ/EMBL/GenBank databases. Takeshi K Watanabe, Otsuka Pharmaceutical Co., Ltd., Otsuka GEN Research Institute, 463-10, Kagasuno, Kawachi-cho, Tokushima, Tokushima 771-0192, Japan (E-mail:watanabe@otsuka-genome.ad.jp, Tel:81-886-65-2888, Fax:81-886-37-1035)
FEATURES
source Location/Qualifiers
1..281
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="OT58.09/882h07"
/note="OT58.09/882h07F=5'-CGATCCATTCGCGGTCGTC-3', OT58.09/882h07R=5'-GTCACTCTCACTCCACAGTG-3'"

BASE COUNT 42 a 42 c 71 g 126 t

Query Match 100.0%; Score 21; DB 11; Length 281;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 128 AGAACAACAACAACAACA 108

RESULT 7
HUMRPTPLM 309 bp DNA linear PRI 11-MAR-1996
LOCUS
DEFINITION Human microsatellite dinucleotide repeat polymorphism.
ACCESSION L20486
VERSION L20486.1 GI:1220359
KEYWORDS microsatellite polymorphism; microsatellite repeat; repeat polymorphism.
SOURCE Homo sapiens (tissue library: Bluescript) DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (sites)
AUTHORS Sainz,J. and Pulst,S.-M.
JOURNAL Unpublished (1993)
FEATURES
source Location/Qualifiers
1..309
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="22q12"
/tissue_lib="Bluescript"
212..248
/note="polymorphism"
/rpt_family="microsatellite"
/rpt_type=tandem

repeat_region

BASE COUNT 113 a 82 c 52 g 62 t

Query Match 100.0%; Score 21; DB 9; Length 309;
Best Local Similarity 100.0%; Pred. No. 2,2e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaaacaacaacaaca 21
|||||

Db 41 AGAAACAACAACAACA 61

RESULT 8
HUMUT925 381 bp DNA linear SNS 26-JUL-1993

LOCUS Human chromosome 10 STS UT925, sequence tagged site.
DEFINITION L16357
VERSION L16357.1 GI:308719
KEYWORDS STS; PCR primer; microsatellite marker; microsatellite repeat;
repeat polymorphism; sequence tagged site; tetranucleotide repeat.
Homo sapiens DNA.

SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
Gerken, S.C., Matsunami, N., Lawrence, E., Carlson, M., Moore, M.,
1 (bases 1 to 381)
Ballard, L., Melis, R., Robertson, M., Bradley, P., Elsnier, T.,
Tingey, A., Rodriguez, P., Albertsen, H., Lalouel, J.-M. and White, R.
Genetic and physical mapping of simple sequence repeat containing
sequence tagged sites from the human genome
Unpublished (1993)

JOURNAL Submitted by: Utah Center for Human Genome Research University of
Utah, Dept. of Human Genetics
2160 Eccles Institute of Human Genetics
Salt Lake City, UT 84112
e-mail: sts@corona.med.utah.edu
Primer A: GCACCTCCACCTATGCAAC
Primer B: GTCCCTGTGAGAACTGGATG
32P-label: A Primer
PCR Profile:
Initial Denaturation: 94C 300sec
PCR Cycles: 5
Denaturation: 94C 10sec
Annealing: 60C 10sec
Extension: 72C 20sec
Mg++: 1mM
Gel: Acrylamide 7%, Formamide 32%, Urea 34%
Alleles: 7.

FEATURES
source Location/Qualifiers
1..381
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="10"
STS
49..223
/standard_name="STS UT925"
primer_bind 49..67
Homo sapiens (D8S320 locus) DNA sequence, tetranucleotide repeat
polymorphism.
BASE COUNT 156 a 82 c 94 g 48 t 1 others
ORIGIN

Query Match 100.0%; Score 21; DB 11; Length 381;
Best Local Similarity 100.0%; Pred. No. 2,2e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaaacaacaacaaca 21
|||||

Db 171 AGAAACAACAACAACA 191

RESULT 9
HUMUT1400 407 bp DNA linear STS 26-JUL-1993

LOCUS Human STS UT1400, sequence tagged site.
DEFINITION

ACCESSION L17810
VERSION L17810.1 GI:307674
KEYWORDS STS; PCR primer; STS sequence; microsatellite marker;
microsatellite repeat; repeat polymorphism; sequence tagged site.
Homo sapiens DNA.

SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 407)
Ballard, L., Melis, R., Robertson, M., Bradley, P., Elsnier, T.,
Tingey, A., Rodriguez, P., Albertsen, H., Lalouel, J.-M. and White, R.
Genetic and physical mapping of simple sequence repeat containing
sequence tagged sites from the human genome
Unpublished (1993)

JOURNAL Submitted by: Utah Center for Human Genome Research University of
Utah, Dept. of Human Genetics
2160 Eccles Institute of Human Genetics
Salt Lake City, UT 84112
e-mail: sts@corona.med.utah.edu
Primer A: GATGGAGAGCTGCTTAAG
Primer B: TGGCATGCTCCTGAGTAG
32P-label: B Primer
PCR Profile:
Initial Denaturation: 94C 300sec
PCR Cycles: 5
Denaturation: 94C 10sec
Annealing: 62C 10sec
Extension: 72C 20sec
Mg++: 1mM
Gel: Acrylamide 7%, Formamide 32%, Urea 34%
Alleles: 1.

FEATURES
source Location/Qualifiers
1..407
/organism="Homo sapiens"
/db_xref="taxon:9606"
STS
68..362
/standard_name="STS UT1400"
primer_bind 68..87
Homo sapiens (D8S320 locus) DNA sequence, tetranucleotide repeat
polymorphism.
BASE COUNT 166 a 63 c 109 g 61 t 8 others
ORIGIN

Query Match 100.0%; Score 21; DB 11; Length 407;
Best Local Similarity 100.0%; Pred. No. 2,2e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaaacaacaacaaca 21
|||||

Db 273 AGAAACAACAACAACA 293

RESULT 10
HUMDBS320T 497 bp DNA linear PRI 08-OCT-1993

LOCUS Homo sapiens (D8S320 locus) DNA sequence, tetranucleotide repeat
polymorphism.
DEFINITION L12265
VERSION L12265.1 GI:388013
KEYWORDS tetranucleotide repeat polymorphism.
Homo sapiens DNA.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 497)
Riley, R., Nelson, L., Lu, J., Robertson, M., Ballard, L., Connolly, J.
and Ward, K.
Tetranucleotide repeat polymorphism at the D8S320 locus
Hum. Mol. Genet. 2, 1512-1512 (1993)
Location/Qualifiers
1..497
/organism="Homo sapiens"

/isolate="KW205"
/db_xref="taxon:9606"
/clone_lib="flow-sorted chromosome 8 specific cosmid
library (LA08MC01)"
repeat_region 1..497
/standard_name="D8S320 locus"
/rpl_unit="AAG

BASE COUNT 270 a 45 c 105 g 62 t 15 others
ORIGIN

Query Match 100.0%; Score 21; DB 9; Length 497;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaacaacaacaacaaca 21
|||||

Db 287 AGAACAACAACAACAACA 307

RESULT 11
HUMUT5307 497 bp DNA linear STS 26-JUL-1993
LOCUS Human chromosome 8 STS UT5307, sequence tagged site.
ACCESSION U18590.1 GI:308301
VERSION U18590.1 GI:308301
KEYWORDS STS; PCR primer; STS sequence; microsatellite marker;
microsatellite repeat; repeat polymorphism; sequence tagged site.
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 497)
Gerken, S.C., Matsunami, N., Lawrence, E., Carlson, M., Moore, M.,
Ballard, L., Mellis, R., Robertson, M., Bradley, P., Eisner, T.,
Tingey, A., Rodriguez, P., Albertsen, H., Lalouel, J.-M. and White, R.
Genetic and physical mapping of simple sequence repeat containing
sequence tagged sites from the human genome
Unpublished (1993)
Submitted by: Utah Center for Human Genome Research University of
Utah, Dept. of Human Genetics
2160 Eccles Institute of Human Genetics
Salt Lake City, UT 84112
e-mail: sts@corona.med.utah.edu
Primer A: GGATTAAGCTGCAATGCAAAAC
Primer B: GTCAATTCAGAGACTGTAAACAG
PCR Profile:
Initial Denaturation: 94C 300sec
PCR Cycles: 5
Denaturation: 94C 10sec
Annealing: 60C 10sec
Extension: 72C 20sec
Mg++: 2mM
Gel: Acrylamide 7%, Formamide 32%, Urea 34%
Alleles: 4.

FEATURES
source 1.497
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="8"
STS 38..449
/standard_name="STS UT5307"
primer_bind 38..62
primer_bind complement(425..449)
BASE COUNT 270 a 45 c 105 g 62 t 15 others
ORIGIN

Query Match 100.0%; Score 21; DB 11; Length 497;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaacaacaacaacaaca 21

Db 287 AGAACAACAACAACAACA 307
|||||

RESULT 12
G24380/c 503 bp DNA linear STS 31-MAY-1996
LOCUS human STS WI-31303, sequence tagged site.
ACCESSION G24380
VERSION G24380.1 GI:1344706
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE human STS derived from sequences in dbEST and the UniGene
collection.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 503)
Hudson, T.
Whitehead Institute/MIT Center for Genome Research; Physically
Mapped STS
Unpublished

JOURNAL
COMMENT
Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Whitehead Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu

Primer A: AAGCCCTGTAACCTGAGCT
Primer B: GGCAGAACATCTCGAATA
STS size: 150
PCR Profile:
Presoak:
Denaturation:
Annealing: 56 degrees C
Polymerization:
PCR Cycles: 35
Thermal Cycler:
Protocol:
Template: 10 ng
Primer: each 5 pM
dNTPs: each 4 mM
Tag Polymerase: 0.025 units/ul
Total Vol: 20 ul

Buffer:
MgCl2: 1.5 mM
KCl: 50 mM
Tris-HCL: 10 mM
pH: 9.3

Derived from dbEST (genbank accession H92026).
Location/Qualifiers
1..503
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="183.2 cR from top of Chr7 linkage group"
STS 1..150
primer_bind 1..21
primer_bind complement(131..150)
BASE COUNT 145 a 97 c 118 g 128 t 15 others
ORIGIN

Query Match 100.0%; Score 21; DB 11; Length 503;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaacaacaacaacaaca 21
|||||

Db 181 AGAACAACAACAACAACA 161

RESULT 13
LOCUS G42246/c 552 bp DNA linear STS 29-NOV-1999
DEFINITION 58p18c1 Human PAC clone Homo sapiens STS genomic, sequence tagged site.
ACCESSION G42246
VERSION G42246.1 GI:4115876
KEYWORDS STS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 552)
AUTHORS Paavola,P., Avela,K., Horelli-Kuitunen,N., Barlund,M., Kallioniem,A., Idenheimo,N., Kyttälä,M., de la Chapelle,A., Palotie,A., Lehesjoki,A.-E. and Peltonen,L.
High-resolution physical and genetic mapping of the critical region for Meckel syndrome and Mulibrey Nanism on chromosome 17q22-q23
Genome Res. 9 (3), 267-276 (1999)
JOURNAL 99177349
MEDLINE 10077533
COMMENT Contact: Kristina Avela
Department of Medical Genetics
University of Helsinki
Helsinki, Finland
Email: Kristina.Avela@helsinki.fi
Primer A: TATGTCAGATGTCAGTCG
Primer B: CCGCATTTATCTGACAAAGCC
STS size: 107.
Location/Qualifiers
1..552
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Human PAC clone"
/lab_host="E. coli DH10B"
/note="vector: pCYPAC-1; V-type: pl-derived vector"
11..217
111..132
111..217
STS primer_bind complement(196..217)
BASE COUNT 152 a 107 c 127 g 165 t 1 others
ORIGIN

Query Match 100.0%; Score 21; DB 11; Length 552;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 agaaacaacaacaacaacaa 21
|||||
Db 338 AGAACAACAACAACAACA 318

RESULT 14
LOCUS HUMUT6478 561 bp DNA linear STS 29-DEC-1994
DEFINITION Human STS UT6478, sequence tagged site.
ACCESSION L30582
VERSION L30582.1 GI:604976
KEYWORDS STS; PCR primer; STS sequence; microsatellite DNA; microsatellite marker; sequence tagged site; tetranucleotide repeat.
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 561)
AUTHORS Getken,S.C., Matsunami,N., Plaetke,R., Albertsen,H., Ballard,L., Melis,R., Lawrence,E., Moore,M., Holik,P.R., Carlson,M., Zhao,X., Robertson,M., Bradley,P., Elsner,T., Tingey,A., Lalouel,J.-M. and White,R.
Genetic and physical mapping of simple sequence repeat containing

JOURNAL Unpublished (1994)
COMMENT Submitted by: Utah Center for Human Genome Research University of Utah, Dept. of Human Genetics
2160 Eccles Institute of Human Genetics
Salt Lake City, UT 84112
e-mail: sts@corona.med.utah.edu
Primer A: TACATCCAGCTGCGCAC
Primer B: CTCCTCAGAGTCTCCATCG
End to Label: Primer B
PCR Profile:
Initial Denaturation: 94C 30sec
Cycles Denaturation Annealing Extension 5 94
C 10 sec. 64 C 10 sec. 72 C 20 sec. 30
60 C 10 sec. 72 C 20 sec. Mg++: 1.00 mM
Gel: Acrylamide 7%, Formamide 32%, Urea 34%
Alleles: 9.

FEATURES
source Location/Qualifiers
1..561
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="2"
167..343
/standard_name="STS UT6478"
167..186
/evidence=experimental
complement(324..343)
/evidence=experimental

BASE COUNT 170 a 153 c 135 g 93 t 10 others
ORIGIN

Query Match 100.0%; Score 21; DB 11; Length 561;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 agaaacaacaacaacaacaa 21
|||||
Db 269 AGAACAACAACAACAACA 289

RESULT 15
LOCUS AF254660/c 617 bp DNA linear VRT 02-MAY-2001
DEFINITION Tetrao urogallus microsatellite TUT8 sequence.
ACCESSION AF254660
VERSION AF254660.1 GI:13926063
KEYWORDS western capercaillie.
SOURCE Tetrao urogallus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauia; Aves; Neognathae; Galliformes; Tetraonidae; Tetrao.

REFERENCE 1 (bases 1 to 617)
AUTHORS Segelbacher,G., Paxton,R. and Steinbreck,G.
Characterization of hypervariable microsatellites in Capercaillie (Tetrao urogallus)
Unpublished

JOURNAL 2 (bases 1 to 617)
AUTHORS Segelbacher,G.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) Unit of Wildlife Research and Management, TU Munich, Am Hochanger 13, Freising 85354, Germany

FEATURES
source Location/Qualifiers
1..617
/organism="Tetrao urogallus"
/db_xref="taxon:100830"
1..617
/note="microsatellite TUT8"
repeat_region /rpt_type=tandem

BASE COUNT 62 a 156 c 69 g 330 t
ORIGIN

Thu Aug 1 08:29:59 2002

us-09-899-718a-3.rge

Page 7

Query Match 100.0%; Score 21; DB 5; Length 617;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 agaaacaaacaaacaaacaa 21
|||||
Db 58 AGAAACAAACAAACAAACAA 38

Search completed: July 31, 2002, 14:00:51
Job time: 17493 sec

Db 2363 CGCAATTCACTGCGCCAGCAGC 2387

RESULT 2

US-08-162-809-13
; Sequence 13, Application US/08162809
; Patent No. 5457048
; GENERAL INFORMATION:
; APPLICANT: Pasquale, Elena B.
; APPLICANT: Sajjadi, Fereydoon G.
; TITLE OF INVENTION: NOVEL EPH-RELATED TYROSINE KINASES,
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES, AND METHODS OF USE
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States of America
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/162,809
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 9503
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3591 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..2965
; US-08-162-809-13

Query Match 64.1%; Score 18.6; DB 1; Length 3591;
Best Local Similarity 84.0%; Pred. No. 18;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 cgcgaagcgctgcgcgcagcagc 27
||||| ||||||| |||
Db 2408 CGCAATTCACTGCGCCAGCAGC 2432

RESULT 3
US-09-095-443-1
; Sequence 1, Application US/09095443
; Patent No. 6342593
; GENERAL INFORMATION:
; APPLICANT: Plowman, Gregory
; APPLICANT: Peles, Eliot
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT
; TITLE OF INVENTION: OF ALP RELATED DISORDERS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles

STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: FastSeq for Windows 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/095,443
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/049,477
FILING DATE: June 12, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 235/055
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4456 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-095-443-1

Query Match 62.1%; Score 18; DB 4; Length 4456;
Best Local Similarity 80.8%; Pred. No. 31;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 gcaaaagcgctgcgcgcagcagc 29
||||| ||||||| |||
Db 1846 GCCATGCGGTTGCGCCAGCAGC 1871

RESULT 4
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 60.0%; Score 17.4; DB 4; Length 4403765;
Best Local Similarity 75.9%; Pred. No. 38;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;


```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/325,630
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/956,993
FILING DATE: 06-OCT-1992
ATTORNEY/AGENT INFORMATION:
NAME: Oblon, No. 575031man F.
REGISTRATION NUMBER: 24,618
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1207 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Hepatitis C virus
IMMEDIATE SOURCE:
CLONE: J1-1325
FEATURE:
NAME/KEY: CDS
LOCATION: 2..1207
US-08-325-630-1
```

```

Query Match          57.9%; Score 16.8; DB 1; Length 1207;
Best Local Similarity 75.0%; Pred. No. 96;
Matches 21; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
```

```

QY 1 cagcgaaagcgctgcgcagccacaga 28
    |||||
DB 1185 CAGGCAAGACGCGTGCAGCCAGCA 1158
```

```

RESULT 11
US-09-105-537-30/c
Sequence 30, Application US/09105537A
Patent No. 6265202
GENERAL INFORMATION:
APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600,438US1
CURRENT APPLICATION NUMBER: US/09/105,537A
CURRENT FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 30
LENGTH: 13842
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-105-537-30
```

```

Query Match          57.9%; Score 16.8; DB 4; Length 13842;
Best Local Similarity 90.0%; Pred. No. 1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```

QY 9 ggcgcgtgcgcagccacaga 28
    |||||
DB 11075 GGCgcgtgcgcagccacaga 11056
```

```

RESULT 12
US-09-105-537-5/c
Sequence 5, Application US/09105537A
Patent No. 6265202
GENERAL INFORMATION:
APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600,438US1
CURRENT APPLICATION NUMBER: US/09/105,537A
CURRENT FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 5
LENGTH: 36778
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-105-537-5
```

```

Query Match          57.9%; Score 16.8; DB 4; Length 36778;
Best Local Similarity 90.0%; Pred. No. 1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```

QY 9 ggcgcgtgcgcagccacaga 28
    |||||
DB 12816 GGCgcgtgcgcagccacaga 12797
```

```

RESULT 13
US-09-320-878-19/c
Sequence 19, Application US/09320878A
Patent No. 6117659
GENERAL INFORMATION:
APPLICANT: ASHLEY, Gary
APPLICANT: BETLACH, Melanie C.
APPLICANT: BETLACH, Mary C.
APPLICANT: MCDANIEL, Robert
APPLICANT: TANG, Li
TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
FILE REFERENCE: 300622002120
CURRENT APPLICATION NUMBER: US/09/320,878A
CURRENT FILING DATE: 1998-05-27
EARLIER APPLICATION NUMBER: CIP OF 09/141,908
EARLIER FILING DATE: 1998-08-28
EARLIER APPLICATION NUMBER: CIP OF 09/073,538
EARLIER FILING DATE: 1998-05-06
EARLIER APPLICATION NUMBER: CIP OF 08/846,247
EARLIER FILING DATE: 1997-04-30
EARLIER APPLICATION NUMBER: 60/119,139
EARLIER FILING DATE: 1999-02-08
EARLIER APPLICATION NUMBER: 60/100,880
EARLIER FILING DATE: 1998-09-22
EARLIER APPLICATION NUMBER: 60/087,080
EARLIER FILING DATE: 1998-05-28
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 19
LENGTH: 38506
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-320-878-19
```

```

Query Match          57.9%; Score 16.8; DB 3; Length 38506;
Best Local Similarity 90.0%; Pred. No. 1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```

QY 9 ggcgcgtgcgcagccacaga 28
    |||||
DB 10958 GGCgcgtgcgcagccacaga 10939
```

RESULT 14
US-08-913-362-3
; Sequence 3, Application US/08913362
; Patent No. 6287574
GENERAL INFORMATION:
APPLICANT: Brodeur, Bernard R
APPLICANT: Martlin, Denis
APPLICANT: Hamel, Josee
APPLICANT: Rioux, Clement
TITLE OF INVENTION: PROTEINASE K RESISTANT SURFACE PROTEIN
TITLE OF INVENTION: OF NEISSERIA MENINGITIDIS
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/913,362
FILING DATE: 13-NOV-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/406,362
FILING DATE: 17-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/001,983
FILING DATE: 04-AUG-1995
ATTORNEY/AGENT INFORMATION:
NAME: Bent, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 047998/0128
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 710 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Neisseria meningitidis
STRAIN: MCH88
FEATURE:
NAME/KEY: CDS
LOCATION: 116..643
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 116..172
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 173..643
US-08-913-362-3

Query Match 56.6%; Score 16.4; DB 4; Length 710;
Best Local Similarity 76.9%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 4 gcaagcgctgcgcgcacgcacgac 29
||| ||||||||| |||||

Db 363 GCATCGCGCGCTCCGCCATTACGAC 388
RESULT 15
US-08-913-362-1
; Sequence 1, Application US/08913362
; Patent No. 6287574
GENERAL INFORMATION:
APPLICANT: Brodeur, Bernard R
APPLICANT: Martlin, Denis
APPLICANT: Hamel, Josee
APPLICANT: Rioux, Clement
TITLE OF INVENTION: PROTEINASE K RESISTANT SURFACE PROTEIN
TITLE OF INVENTION: OF NEISSERIA MENINGITIDIS
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/913,362
FILING DATE: 13-NOV-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/406,362
FILING DATE: 17-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/001,983
FILING DATE: 04-AUG-1995
ATTORNEY/AGENT INFORMATION:
NAME: Bent, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 047998/0128
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 830 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Neisseria meningitidis
STRAIN: 608B
FEATURE:
NAME/KEY: CDS
LOCATION: 143..667
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 143..199
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 200..667
US-08-913-362-1

Query Match 56.6%; Score 16.4; DB 4; Length 830;
Best Local Similarity 76.9%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 4 gcaagcgctgcgcgcacgcacgac 29

Thu Aug 1 08:29:58 2002

us-09-899-718a-2.rni

Page 7

Db 387 GCATCGGCGCGCTCCGCCATTACGAC 412

Search completed: July 31, 2002, 12:20:34
Job time: 11657 sec

1
:
.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 12:11:51 ; Search time 5855.71 Seconds
(without alignments)
66.843 Million cell updates/sec

Title: US-09-899-718a-2

Perfect score: 29
Sequence: 1 caccgcaagcgccgctgcgcacgacacac 29

Scoring table: IDENTITY_NTC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estdb:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_estl:*
10: gb_est2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	19.8	68.3	222	10	N46230 YY72B11.r1
C 2	19.8	68.3	255	9	AA227176 ZR20G09.r
C 3	19.8	68.3	350	10	W94459 ZEL2E03.r1
C 4	19.8	68.3	984	10	BG024310 602274479
C 5	19.8	68.3	2306	11	BC004149 Homo sapi
C 6	19.4	66.9	1028	10	BE733309 601569494
C 7	19.4	66.9	1100	10	BE893877 601436242
C 8	19.4	66.9	1829	12	AO782504 HS.3174.A
C 9	19.2	66.2	325	10	DA8137 R1C514202A
C 10	19.2	66.2	1096	12	CNS04XLD
C 11	19.2	65.5	290	9	AM972385 EST384476
C 12	19.2	65.5	326	9	AV433974 AV433974
C 13	19.2	65.5	416	9	AV435705 AV435705
C 14	19.2	65.5	452	9	AV430263 AV430263
C 15	19.2	65.5	479	9	BB855054 BB855054
C 16	19.2	65.5	509	9	AV435479 AV435479
C 17	19.2	65.5	517	9	AV433024 AV433024

C 18	19	65.5	519	9	AV435216	AV435216
C 19	19	65.5	519	9	AV435485	AV435485
C 20	19	65.5	528	9	AV432179	AV432179
C 21	19	65.5	528	9	AV434482	AV434482
C 22	19	65.5	529	9	AV432635	AV432635
C 23	19	65.5	530	9	AV433071	AV433071
C 24	19	65.5	530	9	AV435732	AV435732
C 25	19	65.5	532	9	AV432358	AV432358
C 26	19	65.5	533	9	AV434711	AV434711
C 27	19	65.5	534	9	AV432450	AV432450
C 28	19	65.5	534	9	AV433137	AV433137
C 29	19	65.5	534	9	AV434241	AV434241
C 30	19	65.5	536	9	AV435646	AV435646
C 31	19	65.5	544	9	AV434121	AV434121
C 32	19	65.5	546	9	AV432203	AV432203
C 33	19	65.5	548	9	AV434539	AV434539
C 34	19	65.5	549	9	AV432005	AV432005
C 35	19	65.5	549	9	AV437069	AV437069
C 36	19	65.5	552	9	AV431821	AV431821
C 37	19	65.5	552	10	BG274284	BG274284
C 38	19	65.5	571	10	BF258049	BF258049
C 39	19	65.5	789	10	BG716721	BF258049
C 40	19	65.5	917	10	BF213345	BF213345
C 41	19	65.5	1166	11	AK021174	AK021174
C 42	18.6	64.1	301	12	A2922255	A2922255
C 43	18.6	64.1	500	10	BE587380	BE587380
C 44	18.6	64.1	684	12	AQ329925	AQ329925
C 45	18.6	64.1	793	10	BG923682	BG923682

ALIGNMENTS

RESULT 1
LOCUS N46230 222 bp mRNA linear EST 14-FEB-1996
DEFINITION yy72b11.r1 Soares multiple_sclerosis-2NBHMSP Homo sapiens cDNA
clone IMAGE:279069 5', mRNA sequence.

ACCESSION N46230
VERSION N46230.1 GI:1187396

KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 222)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman

Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevas, E., Waterston

, R., Williamson, A., Wohlmann, P. and Wilson, R.

Unpublished (1995)
Contact: Wilson RK

Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800
Fax: 314 286 1810

Email: estevatson.wustl.edu
This clone is available royalty-free through LML ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: T7

High quality sequence stop: 185.
location/Qualifiers

FEATURES

source
1..222
/organism="Homo sapiens"
/db_xref="GDB:3897445"
/db_xref="taxon:9606"
/clone="IMAGE:279069"
/clone_lib="Soares_multiple_sclerosis_2NBHMSP"
/sex="male"
/tissue_type="multiple sclerosis lesions"
/dev_stage="Age 46"
/lab_host="DH10B (ampicillin resistant)"

```

/Note="Vector: pT73D (Pharmacia) with a modified
polylinker V-type: phagemid; Site_1: Not I; Site_2: Eco RI
; 1st strand cDNA was primed with a Not I - oligo(dT)
primer (5'
TCGTACCATCTCGAAGTGGAGCGCGCGCATTTTTTTTTTTTTTTT 3'),
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M. Fatima Bonaldo. RNA from 4 multiple sclerosis
lesions from one patient was kindly provided by Dr. Kevin
G. Becker (NINDS/NIH).
G. Becker (NINDS/NIH).
BASE COUNT      43 a      71 c      64 g      38 t      6 others
ORIGIN

Query Match      68.3%; Score 19.8; DB 10; Length 222;
Best Local Similarity 91.3%; Pred. No. 4.5e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      4      gcaaaagcgcgtcgccagccac 26
          |||||||
Db      151      GCAAAAGCGCGTCGCGCGGCAC 129

RESULT 2      AA227176      255 bp      mRNA      linear      EST 11-MAR-1998
LOCUS      zr20909.r1 StrataGene NT2 neuronal precursor 937230 Homo sapiens
DEFINITION      CDNA clone IMAGE:664000 5', mRNA sequence.
ACCESSION      AA227176
VERSION      AA227176.1 GI:1848846
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 255)
AUTHORS      Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisler, G., Jost, S.,
Kritzman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin
J., Moore, B., Schellenberg, K., Stepcoe, M., Tan, F., Theisling, B.,
White, Y., Wylie, T., Waterston, R. and Wilson, R.
WashU-NCI human EST Project
Unpublished (1997)
TITLE      JOURNAL
COMMENT      Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1138 Std Error: 0.00
Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 247.
Location/Qualifiers
1. 255
/organism="Homo sapiens"
/db_xref="GDB:5425859"
/db_xref="taxon:9606"
/clone_lib="StrataGene NT2 neuronal precursor 937230"
/tissue_type="neuroepithelial cells"
/dev_stage="Ntera-2 neuroepithelial cells"
/lab_host="SOLR (kanamycin resistant)"
/Note="Organ: brain; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
oligo dT. Undigested, exponentially growing neuroepithelial
cells (Ntera-2/cl.D1). Average insert size: 1.0 kb;
Uni-TAP XR Vector; -5' adaptor sequence: 5' GAATTTCGGCAGCAG
3' -3' adaptor sequence: 5' CTCGAGTTTGTTTTGTTTT 3'."

BASE COUNT      44 a      71 c      75 g      65 t

/Note="Vector: pT73D (Pharmacia) with a modified
polylinker V-type: phagemid; Site_1: Not I; Site_2: Eco RI
; 1st strand cDNA was primed with a Not I - oligo(dT)
primer (5'
TCGTACCATCTCGAAGTGGAGCGCGCATTTTTTTTTTTTTTTT 3'),
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M. Fatima Bonaldo. RNA from 4 multiple sclerosis
lesions from one patient was kindly provided by Dr. Kevin
G. Becker (NINDS/NIH).
G. Becker (NINDS/NIH).
BASE COUNT      43 a      71 c      64 g      38 t      6 others
ORIGIN

Query Match      68.3%; Score 19.8; DB 9; Length 255;
Best Local Similarity 91.3%; Pred. No. 4.7e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      4      gcaaaagcgcgtcgccagccac 26
          |||||||
Db      97      GCAAAAGCGCGTCGCGCGGCAC 75

RESULT 3      W94459      350 bp      mRNA      linear      EST 16-JUL-1996
LOCUS      zel2e03.r1 Soares fetal heart.NbHH19W Homo sapiens CDNA clone
DEFINITION      IMAGE:358780 5', mRNA sequence.
ACCESSION      W94459
VERSION      W94459.1 GI:1423589
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 350)
AUTHORS      Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman
M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaaskis, E., Waterston
R., Williamson, A., Wohlmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
TITLE      JOURNAL
COMMENT      Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: mob. REGA+ET
High quality sequence stop: 321.
Location/Qualifiers
1. 350
/organism="Homo sapiens"
/db_xref="GDB:1275324"
/db_xref="taxon:9606"
/clone_lib="Soares_fetal_heart_NbHH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/Note="Organ: heart; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer (5'
TCGTACCATCTCGAAGTGGAGCGCGCATTTTTTTTTTTTTTTT 3'),
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M. Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NbHH19W."

BASE COUNT      65 a      107 c      100 g      76 t      2 others
ORIGIN

Query Match      68.3%; Score 19.8; DB 10; Length 350;
Best Local Similarity 91.3%; Pred. No. 5.1e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      4      gcaaaagcgcgtcgccagccac 26
          |||||||

```

Db 195 GCAAGGCGCGTCGGCGGCGAC 173

RESULT 4
LOCUS BG024310/c
DEFINITION 602274479F1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4362450 5', mRNA sequence.
ACCESSION BG024310
VERSION BG024310.1 GI:12409756
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: Louis Staudt, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LHAM10006 row: n column: 19

FEATURES
source
Location/Qualifiers
1..984
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4362450"
/clone_lib="NIH_MGC_85"
/tissue_type="Lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: PCMV-SPORE6; Site:1: NotI; Site:2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.867 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC library."

BASE COUNT 192 a 356 c 249 g 187 t

ORIGIN

Query Match 68.3%; Score 19.8; DB 10; Length 984;
Best Local Similarity 91.3%; Pred. No. 6.9e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 gcaaaagcgctgcggccagccac 26
|||||
Db 152 GCAAGGCGCGTCGGCGGCGAC 130

RESULT 5
LOCUS BC004149/c
DEFINITION BC004149 Homo sapiens, clone IMAGE:2819393, mRNA.
ACCESSION BC004149
VERSION BC004149.1 GI:14708750
KEYWORDS HTC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (01-MAR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

REMARK
COMMENT USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgaabs-remail.nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
Contact: amadan@systemsbiology.org
Anup Madan, Rachel Dickhoff, Jessica Fahay, Stephanie Ford, Julia Greene, Mark Ketteman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 2 Row: 0 Column: 3
This clone has the following problem: frame shifted.

FEATURES
source
Location/Qualifiers
1..2306
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2819393"
/tissue_type="Lung, small cell carcinoma"
/clone_lib="NIH_MGC_7"
/lab_host="DH10B-R"
/note="Vector: pOTB7"

BASE COUNT 460 a 764 c 640 g 442 t

ORIGIN

Query Match 68.3%; Score 19.8; DB 11; Length 2306;
Best Local Similarity 91.3%; Pred. No. 8.8e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 gcaaaagcgctgcggccagccac 26
|||||
Db 1354 GCAAGGCGCGTCGGCGGCGAC 1332

RESULT 6
LOCUS BE733309
DEFINITION BE733309 1028 bp mRNA linear EST 15-SEP-2000
ACCESSION BE733309
VERSION BE733309.1 GI:10147301
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LCM540 row: k column: 23
High quality sequence stop: 136.

FEATURES
source
Location/Qualifiers
1..1028
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3844366"
/clone_lib="NIH_MGC_21"
/tissue_type="Choriocarcinoma"
/lab_host="DH10B (phage-resistant)"

JOURNAL DNA Res. 7, 223-227 (2000)
MEDLINE 20363100
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
source
1. 326
/organism="Porphyra yezoensis"
/strain="TU-1"
/db_xref="taxon:2788"
/clone="PM036f02.r"
/clone_lib="Porphyra yezoensis TU-1"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 25 a 147 c 78 g 76 t
ORIGIN

Query Match 65.5%; Score 19; DB 9; Length 326;
Best Local Similarity 81.5%; Pred. No. 1.le+03;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 cgcgaagcgctgcgcgcgcgcgcac 29
|||||
Db 264 CGCACAGCGCGACGCGACCGACGAC 238

RESULT 13
AV435705 416 bp mRNA linear EST 23-AUG-2000
LOCUS AV435705 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone
DEFINITION PM061c04_r 5', mRNA sequence.
ACCESSION AV435705
VERSION AV435705.1 GI:8590930
KEYWORDS EST.
SOURCE Porphyra yezoensis.
ORGANISM Porphyra yezoensis.
Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
Porphyra.
REFERENCE 1 (bases 1 to 416)
AUTHORS Nikaido, I., Asamizu, E., Nakajima, M., Nakamura, Y., Saga, N. and Tabata, S.
TITLE Generation of 10,154 expressed sequence tags from a leafy gametophyte of a marine red alga, Porphyra yezoensis
JOURNAL DNA Res. 7, 223-227 (2000)
MEDLINE 20363100
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
source
1. 416
/organism="Porphyra yezoensis"
/strain="TU-1"
/db_xref="taxon:2788"
/clone="PM061c04_r"
/clone_lib="Porphyra yezoensis TU-1"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 47 a 178 c 104 g 87 t
ORIGIN

Query Match 65.5%; Score 19; DB 9; Length 416;
Best Local Similarity 81.5%; Pred. No. 1.le+03;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 cgcgaagcgctgcgcgcgcgcgcac 29
|||||
Db 276 CGCACAGCGCGACGCGACCGACGAC 250

RESULT 14
AV430263 452 bp mRNA linear EST 23-AUG-2000
LOCUS AV430263 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone
DEFINITION PL015h12_r 5', mRNA sequence.
ACCESSION AV430263
VERSION AV430263.1 GI:8585488
KEYWORDS EST.
SOURCE Porphyra yezoensis.
ORGANISM Porphyra yezoensis.
Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
Porphyra.
REFERENCE 1 (bases 1 to 452)
AUTHORS Nikaido, I., Asamizu, E., Nakajima, M., Nakamura, Y., Saga, N. and Tabata, S.
TITLE Generation of 10,154 expressed sequence tags from a leafy gametophyte of a marine red alga, Porphyra yezoensis
JOURNAL DNA Res. 7, 223-227 (2000)
MEDLINE 20363100
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
source
1. 452
/organism="Porphyra yezoensis"
/strain="TU-1"
/db_xref="taxon:2788"
/clone="PL015h12_r"
/clone_lib="Porphyra yezoensis TU-1"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 48 a 186 c 119 g 99 t
ORIGIN

Query Match 65.5%; Score 19; DB 9; Length 452;
Best Local Similarity 81.5%; Pred. No. 1.le+03;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 cgcgaagcgctgcgcgcgcgcgcac 29
|||||
Db 267 CGCACAGCGCGACGCGACCGACGAC 241

RESULT 15
BB855054 479 bp mRNA linear EST 26-NOV-2001
LOCUS BB855054 RIKEN full-length enriched, B16 F10Y cells Mus musculus
DEFINITION cDNA clone G370022K07 5', mRNA sequence.
ACCESSION BB855054
VERSION BB855054.1 GI:17096508
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Euteleostomi; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 479)
AUTHORS Akimura, T., Arikawa, T., Carninci, P., Furuno, M., Hanagaki, T., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Imotani, K., Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takehashi, F., Takaku-Akahita, S., Tanaka, T., Tomaru, A., Toyota, T., Wachihi, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
TITLE RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al. 2001)
JOURNAL unpublished (2001)
COMMENT Contact: Yoshinide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsr.riken.go.jp,

URL: <http://genome.gsc.riken.go.jp/>

, Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh

, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

Waghi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,

Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura

, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and

Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara

, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for

further details.

e mouse tissues.

FEATURES

source

location/Qualifiers

1. 479

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="G370022K07"

/clone_lib="RIKEN full-length enriched, B16 F10Y cells"

/cell_type="B16 F10Y cells"

BASE COUNT 112 a 109 c 149 g 109 t

ORIGIN

Query Match

Best Local Similarity 81.5%; Score 19; DB 9; Length 479;

Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 cagcgaagcgctcgccagccagc 27

DB 45 CGCGGTAGCGCGCGCGCGCGCG 19

Search completed: July 31, 2002, 12:11:53
Job time: 13466 sec

Query Match 67.68; Score 19.6; DB 22; Length 2473;
Best Local Similarity 84.68; Pred. No. 44;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 caccgaagcgcgtcgcgcagccac 26
|||||
Db 1678 CACGCAAGGCCCTCAGCTTGCCAC 1653

RESULT 2
AAL05201/c
ID AAL05201 standard; DNA: 2474 BP.

AC AAL05201;
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 7889.
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN W0200155320-A2.
PD
XX 02-AUG-2001.
PF
XX 17-JAN-2001; 2001WO-US01339.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225457.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 25-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 26-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239933.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241121.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246539.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.

XX 08-JUN-1995.
 PD
 XX
 PF 07-SEP-1994; 94WO-US10140.
 XX
 PR 03-DEC-1993; 93US-0162809.
 XX
 PA (LJOL-) LA JOLLA CANCER RES FOUND.
 XX
 PI Pasquale EB, Sajjadi FG;
 XX
 DR WPI: 1995-215256/28.
 DR P-PSDB; AAR75708.
 XX
 PT Eph-related protein tyrosine kinase(s) - for monitoring and diagnosing
 PT cancer.
 XX
 PS Claim 2; Page 63-67; 129pp; English.
 XX
 CC Novel Eph-related PTK cDNA clone Cek10 (given in AAQ90656) and a
 CC variant clone, Cek10+ (AAQ90658), whose product contains a 15-amino
 CC acid insertion in the juxtamembrane domain, were isolated from a
 CC chick embryo library in lambda gt11. Cek10 and Cek10+ may
 CC originate from the same gene by alternative splicing. Cek10
 CC expression was prominent in the kidney, and to a lesser extent in
 CC the lung.
 CC
 SQ Sequence 3546 BP; 742 A; 1100 C; 1031 G; 673 T; 0 other;

Query Match 64.1%; Score 18.6; DB 16; Length 3546;
 Best Local Similarity 84.0%; Pred. No. 1.2e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 OY 3 cgcgaagcggtcgccagccagc 27
 ||||| | ||||| ||||| |||||
 Db 2363 cgcgaattcagctcgccagccagc 2387

RESULT 5
 AAQ90658
 ID AAQ90658 standard; cDNA; 3591 BP.
 XX
 AC AAQ90658;
 XX
 DT 11-NOV-1995 (first entry)
 XX
 DE Eph-related PTK Cek10+ cDNA.
 XX
 KW Cek10+; Eph; protein tyrosine-kinase; PTK; cancer; diagnosis;
 KW prognosis; ss.
 XX
 OS Gallus sp.
 XX
 FH Key Location/Qualifiers
 FT CDS 2..2968
 FT /*tag= a
 FT
 PN WO9515375-A.
 XX
 PD 08-JUN-1995.
 XX
 PF 07-SEP-1994; 94WO-US10140.
 XX
 PR 03-DEC-1993; 93US-0162809.
 XX
 PA (LJOL-) LA JOLLA CANCER RES FOUND.
 XX
 PI Pasquale EB, Sajjadi FG;
 XX
 DR WPI: 1995-215256/28.
 DR P-PSDB; AAR75710.
 XX

PT Eph-related protein tyrosine kinase(s) - for monitoring and diagnosing
 PT cancer.
 XX
 PS Claim 2; Page 78-82; 129pp; English.
 XX
 CC Novel Eph-related PTK cDNA clone Cek10 (given in AAQ90656) and a
 CC variant clone, Cek10+ (AAQ90658), whose product contains a 15-amino
 CC acid insertion in the juxtamembrane domain, were isolated from a
 CC chick embryo library in lambda gt11. Cek10 and Cek10+ may
 CC originate from the same gene by alternative splicing. Cek10
 CC expression was prominent in the kidney, and to a lesser extent in
 CC the lung.
 CC
 SQ Sequence 3591 BP; 753 A; 1112 C; 1042 G; 684 T; 0 other;

Query Match 64.1%; Score 18.6; DB 16; Length 3591;
 Best Local Similarity 84.0%; Pred. No. 1.2e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 OY 3 cgcgaagcggtcgccagccagc 27
 ||||| | ||||| ||||| |||||
 Db 2408 cgcgaattcagctcgccagccagc 2432

RESULT 6
 AAF74776
 ID AAF74776 standard; cDNA; 1511 BP.
 XX
 AC AAF74776;
 XX
 DT 17-MAY-2001 (first entry)
 XX
 DE Human DEC2b encoding cDNA sequence SEQ ID NO:11.
 XX
 KW DEC2a; DEC2b; bHLH type transcription factor; DEC2; DEC1;
 KW basic helix loop helix protein; cell differentiation; proliferation; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 2..1456
 FT /*tag= a
 FT /product= "DEC2b"
 FT /note= "bHLH type transcription factor"
 FT
 PN WO200114551-A1.
 XX
 PD 01-MAR-2001.
 XX
 PF 19-JUN-2000; 2000WO-JP03991.
 XX
 PR 19-AUG-1999; 99JP-0233286.
 XX
 PA (CHUS) CHUGAI SEIYAKU KK.
 XX
 PI Fujimoto K, Shin M, Kato Y;
 XX
 DR WPI: 2001-202935/20.
 DR P-PSDB; AAB70693.
 XX
 PT DEC2 is a basic helix loop helix protein of the DEC family for use in
 PT development of drugs for treatment of disorders of cell differentiation
 PT and proliferation -
 XX
 PS Claim 1; Page 63-68; 83pp; Japanese.
 XX
 CC The present invention describes a basic helix loop helix (bHLH) type
 CC transcription factor designated DEC2. DEC2 can be used as a tool in
 CC the development of drugs for the treatment and prevention of disorders
 CC involving cell differentiation and proliferation. The present sequence
 CC encodes the specifically claimed human DEC2b protein, as given in the
 CC present invention.

Db 896 aggcacagcgcgcgcgagccgcgtc 923

KW human; open reading frame, ORF; detection; cytosolic; hepatocellular; neurotrophic;
 KW vulnervary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;

KM anticonvulsant; osteoporotic; antiarthritic; immunosuppressant; cardiant;
 KM immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KM hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KM antiviral; antibacterial; antifungal; antirheumatic; antihypoid;
 KM antinaeamic; gene therapy; cancer; proliferative disorder; hypertension;
 KM neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KM cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KM cholesterol ester storage; systemic lupus erythematosus; infection;
 KM severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KM allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KM bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KM thrombosis; contraceptive; ss.
 KM
 OS Homo sapiens.
 XX
 XX
 PM WO200058473-A2.
 PD
 PD 05-OCT-2000.
 PF
 PF 31-MAR-2000; 2000WO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 XX
 PA (CURA-) CURAGEN CORP.
 PI
 PI Shinkets RA, Leach M;
 DR WPI: 2000-602362/57.
 DR P-PSDB; AAB43064.
 PT
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 PT
 PS Claim 5; Page 4829-4832; 5507pp; English.
 XX
 XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading Frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulneryary;
 CC antiparasitic; antiparkinsonian; nootropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antihypoid; and antinaeamic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX
 XX Sequence 3810 BP; 730 A; 1397 C; 1015 G; 664 T; 4 other;

	Query Match	62.1%	Score 18;	DB 21;	Length 3610;
	Best Local Similarity	80.8%;	Pred No.	2.1e+02;	
Matches	21;	Conservative	0;	Mismatches	5;
				Indels	0;
Gaps	0;				
OY	4 gcaagcgcgctgcggccacgacgac	29			
Db	1243 gccatgaggcgttcctgcgcacgaccacac	1268			

RESULT	10
AAc81226	
ID	AAc81226 standard; cDNA; 4022 BP.
XX	
AC	AAc81226;
XX	
DT	23-FEB-2001 (first entry)
XX	

DE Human tyrosine phosphatase HD-PTP cDNA fragment cKAL16, SEQ ID NO:4.
 XX
 KW Human; histidine domain-protein tyrosine phosphatase; HD-PTP;
 KW chromosome 3p21.3; gene deletion; tumour suppressor; cytosolic;
 KW lung cancer; tumour; gene therapy; diagnosis; recombinant production
 KW anticancer; ss.

AA	
OS	Homo sapiens.
XX	
PN	WC200063392-A1.
XX	
PD	26-Oct-2000.
xx	

AA
PF 14-APR-2000; 2000WO-JP02455.
XX
PR 16-APR-1999; 99JP-0108842.
XX
PA (KYOW) KYOWA HAKKO KOSYO KK.
PA

AA
PI
XX
XX
WPI: 2000-672740/65.
DR
P-ESDB; AAE29665.
XX

aa	
pt	Human tyrosine phosphatase with oncogenic activity encoded by a gene frequently deleted in lung cancer, useful for treatment and diagnosis of tumors
pt	-
pt	
xx	
ps	Example 1; Page 98-106; 134pp; Japanese.
xy	

The invention relates to a novel human tyrosine phosphatase, histidine domain-protein tyrosine phosphatase (HD-PTP; AAB29661) and to human HD-PTP nucleic acids (AAC81224, AAC81225, AAC81262, AAC81263). The HD-PTP gene is located on chromosome 3p21.3. This region is frequently found to be deleted in lung cancers, and is therefore thought to contain a tumour suppressor gene. The invention also relates to expression vectors and host cells containing human HD-PTP nucleic acids; the recombinant production of HD-PTP; anticancer drugs containing HD-PTP; gene therapy compositions containing DNA encoding HD-PTP; diagnostic reagents containing HD-PTP oligonucleotides; antibodies specific for HD-PTP; and an immunoassay method using HD-PTP-specific antibodies for use in cancer diagnosis and investigation. HD-PTP proteins, nucleic acids and antibodies may be used in the treatment, investigation and diagnosis of cancers, particularly those of the lung. The present sequence represents a human HD-PTP cDNA fragment, cKAL16.

Sequence	4022 BP;	776 A;	1456 C;	1081 G;	709 T;	0 other;
Query Match	62.1%;	Score 18;	DB 21;	Length 4022;		
Best local Similarity	80.8%;	Pred. No. 2.1e+02;				
Matches	21;	Conservative	0;	Mismatches	5;	Indels
					0;	Gaps
						0;

QY 4 gcaagcgcgctcggccagccaacgac 29
 || | ||| ||||| ||||| ||
Db 1471 gcaatbgcggttcggcagccaac 1496

RESULT	11
ID	AAV81748
XX	AAV81748 standard; CDNA; 4456 BP.

AA
AC AAV81748;
XX
DT 10-MAR-1999 (first entry)

```
XX DE Human ALP encoding cDNA.
XX AC
XX KW PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;
XX KW type I receptor serine/threonine kinase; cancer; leukemia; lymphoma;
XX KW neurodegenerative disease; neuronal survival; Alzheimer's disease;
XX KW Parkinson's disease; Huntington's disease; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT CDS 313..4137
XX FT /tag= a
XX PN WO9849317-A2.
XX PD 05-NOV-1998.
XX PE 27-APR-1998; 98WO-US08439.
XX PF 23-OCT-1997; 97US-0063595.
XX PR 28-APR-1997; 97US-0044428.
XX PR 20-MAY-1997; 97US-0047222.
XX PR 11-JUN-1997; 97US-0049477.
XX PR 11-JUN-1997; 97US-0049756.
XX PR 18-JUN-1997; 97US-0049914.
XX PA (SUGEN-) SUGEN INC.
XX PI App H, Clary D, Courtneidge SA, Hui TH, Jallat B;
XX PI Markby D, Onrust S, Peles E, Plowman GD;
XX PI WPI: 1999-009434/01.
XX DR P-PSDB; AAW89253.
XX PT New nucleic acid encoding specific protein tyrosine phosphatases -
XX PT useful for identifying specific modulators for treatment and
XX PT prevention of cancer and neurodegenerative disease
XX PS Claim 2; Page 149-150; 199pp; English.
XX XX The present invention describes isolated, enriched or purified nucleic
XX CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The
XX CC present sequence encodes human ALP. The above proteins, other than
XX CC ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify
XX CC substances that modulate their activity (i.e. agonists and antagonists,
XX CC including NBI) in vivo or in vitro. These substances are used to treat
XX CC or prevent diseases associated with abnormal signal transduction
XX CC pathways that involve the proteins, particularly cancer (e.g. leukemia
XX CC and lymphoma), while modulators of ALK-7 (which is a type I receptor
XX CC serine/threonine kinase) are used to promote neuronal survival.
XX CC particularly for treating Alzheimer's, Parkinson's or Huntington's
XX CC diseases. Nucleic acid fragments of the polynucleotides encoding the
XX CC proteins can be used as probes to identify and clone related sequences;
XX CC to detect protein-encoded RNA; to generate transgenic animals and in
XX CC gene therapy (optionally after mutation). Ab are used to determine the
XX CC proteins.
XX SQ Sequence 4456 BP; 929 A; 1564 C; 1184 G; 779 T; 0 other;
```

```
Query Match 62.1%; Score 18; DB 20; Length 4456;
Best Local Similarity 80.8%; Pred. No. 2.1e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
OY 4 gcaaaagcgctcgccagcagcagc 29
   ||| ||||| ||||| ||||| |||
Db 1846 gccatgcygcttcgcccagccaccac 1871
```

```
RESULT 12
AAC81224
ID AAC81224 standard; cDNA; 5234 BP.
```

```
XX AC AAC81224;
XX XX 23-FEB-2001 (first entry)
XX DT
XX DE Human histidine domain-protein tyrosine phosphatase cDNA, SEQ ID NO:1.
XX KW Human; histidine domain-protein tyrosine phosphatase; HD-PTP;
XX KW chromosome 3p21.3; gene deletion; tumour suppressor; cytostatic;
XX KW lung cancer; tumour; gene therapy; diagnosis; recombinant production;
XX KW anticancer; ss.
XX OS Homo sapiens.
XX FH WO200063392-A1.
XX FT 26-OCT-2000.
XX PD 14-APR-2000; 2000WO-JP02455.
XX PF 16-APR-1999; 99JP-0108842.
XX PR (KYOW ) KYOWA HAKKO KOGYO KK.
XX PA Shintzu K;
XX PI WPI: 2000-672740/65.
XX DR P-PSDB; AAB29661.
XX PT Human tyrosine phosphatase with oncostatic activity encoded by a gene
XX PT frequently deleted in lung cancer, useful for treatment and diagnosis
XX PT of tumors
XX PS Claim 4; Page 58-69; 134pp; Japanese.
XX XX The invention relates to a novel human tyrosine phosphatase, histidine
XX CC domain-protein tyrosine phosphatase (HD-PTP; AAB29661) and to human
XX CC HD-PTP nucleic acids (AAC81224, AAC81225, AAC81262, AAC81263). The
XX CC HD-PTP gene is located on chromosome 3p21.3. This region is frequently
XX CC found to be deleted in lung cancers, and is therefore thought to contain
XX CC a tumour suppressor gene. The invention also relates to expression
XX CC vectors and host cells containing human HD-PTP nucleic acids; the
XX CC recombinant production of HD-PTP; anticancer drugs containing HD-PTP;
XX CC gene therapy compositions containing HD-PTP oligonucleotides; antibodies specific for
XX CC reagents containing HD-PTP oligonucleotides; antibodies specific for
XX CC HD-PTP; and an immunoassay method using HD-PTP-specific antibodies for
XX CC use in cancer diagnosis and investigation. HD-PTP proteins, nucleic acids
XX CC and antibodies may be used in the treatment, investigation and diagnosis
XX CC of cancers, particularly those of the lung. The present sequence
XX CC represents cDNA encoding human HD-PTP.
XX SQ Sequence 5234 BP; 1049 A; 1803 C; 1441 G; 941 T; 0 other;
```

```
Query Match 62.1%; Score 18; DB 21; Length 5234;
Best Local Similarity 80.8%; Pred. No. 2.1e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
OY 4 gcaaaagcgctcgccagcagcagc 29
   ||| ||||| ||||| ||||| |||
Db 2683 gccatgcygcttcgcccagccaccac 2708
```

```
RESULT 13
AAC81227
ID AAC81227 standard; DNA; 5436 BP.
XX AC AAC81227;
XX XX 23-FEB-2001 (first entry)
XX DT
XX DE Human tyrosine phosphatase HD-PTP DNA fragment cKAL11, SEQ ID NO:5.
XX XX
```


PR	29-SEP-2000	2000US-0236368	
PR	29-SEP-2000	2000US-0236369	
PR	29-SEP-2000	2000US-0236370	
PR	02-OCT-2000	2000US-0236372	
PR	02-OCT-2000	2000US-0237037	
PR	02-OCT-2000	2000US-0237038	
PR	02-OCT-2000	2000US-0237039	
PR	02-OCT-2000	2000US-0237040	
PR	13-OCT-2000	2000US-0239935	
PR	13-OCT-2000	2000US-0239937	
PR	20-OCT-2000	2000US-0240960	
PR	20-OCT-2000	2000US-0241221	
PR	20-OCT-2000	2000US-0241785	
PR	20-OCT-2000	2000US-0241786	
PR	20-OCT-2000	2000US-0241787	
PR	20-OCT-2000	2000US-0241808	
PR	20-OCT-2000	2000US-0241809	
PR	01-NOV-2000	2000US-0246182	
PR	08-NOV-2000	2000US-0246474	
PR	08-NOV-2000	2000US-0246475	
PR	08-NOV-2000	2000US-0246476	
PR	08-NOV-2000	2000US-0246477	
PR	08-NOV-2000	2000US-0246478	
PR	08-NOV-2000	2000US-0246523	
PR	08-NOV-2000	2000US-0246524	
PR	08-NOV-2000	2000US-0246525	
PR	08-NOV-2000	2000US-0246526	
PR	08-NOV-2000	2000US-0246527	
PR	08-NOV-2000	2000US-0246528	
PR	08-NOV-2000	2000US-0246532	
PR	08-NOV-2000	2000US-0246610	
PR	08-NOV-2000	2000US-0246610	
PR	08-NOV-2000	2000US-0246611	
PR	17-NOV-2000	2000US-0249207	
PR	17-NOV-2000	2000US-0249208	
PR	17-NOV-2000	2000US-0249209	
PR	17-NOV-2000	2000US-0249210	
PR	17-NOV-2000	2000US-0249211	
PR	17-NOV-2000	2000US-0249212	
PR	17-NOV-2000	2000US-0249213	
PR	17-NOV-2000	2000US-0249214	
PR	17-NOV-2000	2000US-0249215	
PR	17-NOV-2000	2000US-0249216	
PR	17-NOV-2000	2000US-0249217	
PR	17-NOV-2000	2000US-0249218	
PR	17-NOV-2000	2000US-0249245	
PR	17-NOV-2000	2000US-0249245	
PR	17-NOV-2000	2000US-0249264	
PR	17-NOV-2000	2000US-0249265	
PR	17-NOV-2000	2000US-0249297	
PR	17-NOV-2000	2000US-0249309	
PR	17-NOV-2000	2000US-0249300	
PR	17-NOV-2000	2000US-0249301	
PR	17-NOV-2000	2000US-0250391	
PR	05-DEC-2000	2000US-0251030	
PR	05-DEC-2000	2000US-0251088	
PR	05-DEC-2000	2000US-0256119	
PR	06-DEC-2000	2000US-0251479	
PR	08-DEC-2000	2000US-0251856	
PR	08-DEC-2000	2000US-0251868	
PR	08-DEC-2000	2000US-0251869	
PR	08-DEC-2000	2000US-0251989	
PR	08-DEC-2000	2000US-0251990	
PR	11-DEC-2000	2000US-0254097	
PR	05-JAN-2001	2001US-0259678	
PA	(HUMA-)	HUMAN GENOME SCI INC.	
XX			
XX	Rosen CA,	Barrash SC,	Ruben SM,
DR	WPI, 2001-483426/52.		

XX	Nucleic acids encoding human immune/hematopoietic antigen polypeptides,	
PT	useful for preventing, diagnosing and/or treating cancers and	
PT	metastasis -	
XX		
PS	Disclosure; SEQ ID NO 21966; 3071bp + Sequence Listing; English.	
XX		
CC	AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (1)	
CC	amino acid sequences given in AAK82170 to AAK91921. (1) have cytostatic	
CC	activity, and can be used in gene therapy and vaccine production. (1)	
CC	protein, and polynucleotides may be used in the prevention, diagnosis and	
CC	treatment of diseases associated with inappropriate (1) expression. For	
CC	example, they may be used to treat disorders associated with decreased	
CC	expression by rectifying mutations or deletions in a patient's genome	
CC	that affect the activity of (1) by expressing inactive proteins or to	
CC	supplement the patients own production of (1). Additionally, (1)	
CC	polynucleotides may be used to produce the secreted (1), by inserting	
CC	the nucleic acids into a host cell and culturing the cell to express the	
CC	protein. (1) proteins and polynucleotides may be used to prevent,	
CC	diagnose and treat immune/hematopoietic-related diseases, especially	
CC	cancers and cancer metastases of haematopoietic-derived cells. AAK64703	
CC	to AAK87694 represent human immune/haematopoietic antigen genomic	
CC	sequences from the present invention. AAK54942 to AAK54950 and AAK82169	
CC	represent sequences used in the exemplification of the present invention.	
XX		
SO	Sequence 7016 BP; 1339 A; 2079 C; 2321 G; 1277 T; 0 other;	
	Query Match	62.1%; Score 18; DB 22; Length 7016;
	Best Local Similarity	80.8%; Pred. No. 2.1e+02;
	Matches	21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Oy	4 gcaaaagcgcgctggccagccagcagac 29	
Db	3019 GCCATGCGCGGTGCGCCAGCACAC 2994	
RESULT 15		
AAK70464/C		
ID	AAK70464 standard; DNA; 7016 BP.	
XX		
AC	AAK70464;	
XX		
PT	06-NOV-2001 (first entry)	
XX		
DE	Human immune/haematopoietic antigen genomic sequence SEQ ID NO:25276.	
XX		
KW	Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;	
KW	cytostatic; gene therapy; vaccine; metastasis; ds.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200157182-A2.	
XX		
PD	09-AUG-2001.	
XX		
PF	17-JAN-2001; 2001WO-US01354.	
XX		
PR	31-JAN-2000; 2000US-0179065.	
PR	04-FEB-2000; 2000US-0180628.	
PR	24-FEB-2000; 2000US-0184664.	
PR	02-MAR-2000; 2000US-0186350.	
PR	16-MAR-2000; 2000US-0189874.	
PR	17-MAR-2000; 2000US-0190076.	
PR	18-APR-2000; 2000US-0198123.	
PR	19-MAY-2000; 2000US-0205515.	
PR	07-JUN-2000; 2000US-0209467.	
PR	28-JUN-2000; 2000US-0214886.	
PR	30-JUN-2000; 2000US-0215135.	
PR	07-JUL-2000; 2000US-0216647.	
PR	07-JUL-2000; 2000US-0216880.	
PR	11-JUL-2000; 2000US-0217487.	
PR	11-JUL-2000; 2000US-0217496.	

PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0225759.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235634.
PR 27-SEP-2000; 2000US-0235636.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.

PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246539.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251038.
PR 05-DEC-2000; 2000US-0251719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251866.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
useful for preventing, diagnosing and/or treating cancers and
metastasis -

Disclosure: SEQ ID NO 25276; 3071bp + Sequence listing; English.

AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
activity, and can be used in gene therapy and vaccine production. (I)
proteins and polynucleotides may be used in the prevention, diagnosis and
treatment of diseases associated with inappropriate (I) expression. For
example, they may be used to treat disorders associated with decreased
expression by rectifying mutations or deletions in a patient's genome
that affect the activity of (I) by expressing inactive proteins or to
supplement the patients own production of (I). Additionally, (I)
polynucleotides may be used to produce the secreted (I), by inserting
the nucleic acids into a host cell and culturing the cell to express the
protein. (I) proteins and polynucleotides may be used to prevent.

CC diagnosis and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AA64703
CC to AA87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AA54942 to AA54950 and AA862169
CC represent sequences used in the exemplification of the present invention.
CC
XX
SQ Sequence 7016 BP; 1339 A; 2079 C; 2321 G; 1277 T; 0 other;

Query Match	62.1%;	Score 18;	DB 22;	Length 7016;
Best Local Similarity	80.88;	Pred. NO.	2.1e+02;	
Matches	21;	Conservative	0;	Mismatches 5; Indels 0; Gaps 0;
QY	4	gcaaaagcgcgctcgagccagccagac	29	
DB	3019	gccatgctggcggttcggccagccaccac	2994	

Search completed: July 31, 2002, 14:11:09
Job time: 17366 sec

Query Match 100.0%; Score 29; DB 6; Length 29;
 Best Local Similarity 100.0%; Pred. No. 0.84;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caccgaagcgctgcgcacgacac 29
 |||||||||||||||||||||
 DB 1 CACGCAAGCGCGCTCGCCAGCAGCAGC 29

RESULT 2
 LOCUS AX349063 3785 bp DNA linear PAT 06-FEB-2002
 DEFINITION Sequence 1 from Patent WO0202785.
 ACCESSION AX349063
 VERSION AX349063.1 GI:18615098
 KEYWORDS
 SOURCE bread wheat.
 ORGANISM Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Pooidae; Triticeae; Triticum.
 1 (sites)
 Sprunck,S., Kluth,A., Becker,D., Luetticke,S. and Loefer,H.
 Promoters of gene expression in plant caryopses
 Patent: WO 0202785-A 1 10-JAN-2002;
 Aventis CropScience GmbH (DE)
 Location/Qualifiers
 1..3785
 /organism="Triticum aestivum"
 /db_xref="taxon:4565"

BASE COUNT 945 a 980 c 899 g 961 t
 ORIGIN

Query Match 100.0%; Score 29; DB 6; Length 3785;
 Best Local Similarity 100.0%; Pred. No. 0.34;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caccgaagcgctgcgcacgacac 29
 |||||||||||||||||||||
 DB 1736 CACGCAAGCGCGCTCGCCAGCAGCAGC 1764

RESULT 3
 LOCUS NP0637DP 1437 bp DNA linear VRL 03-FEB-1999
 DEFINITION Orygia pseudotsugata nuclear polyhedrosis virus genes for DNA
 polymerase and spheroidin-like protein.
 ACCESSION D13306.1 GI:222200
 VERSION D13306
 KEYWORDS DNA polymerase; spheroidin.
 SOURCE Orygia pseudotsugata nuclear polyhedrosis virus DNA.
 ORGANISM Orygia pseudotsugata single capsid nuclear polyhedrosis virus
 Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
 Nucleopolyhedrovirus.
 1 (bases 1 to 1437)
 Gross,C.H., Wolgamot,G.M., Russell,R.L., Pearson,M.N. and
 Rohmann,G.F.
 A 37-kilodalton glycoprotein from a baculovirus of Orygia
 pseudotsugata is localized to cytoplasmic inclusion bodies
 J. Virol. 67 (1), 469-475 (1993)
 93100831
 COMMENT Submitted (28-SEP-1992) to DDBJ by:
 George F. Rohmann
 Dept. of Agricultural Chemistry
 Oregon State University
 Corvallis, Oregon 97331-6502
 U.S.A.
 Phone: 503-737-1793
 Email: rohmann@crb.orst.edu
 Fax: 503-737-0497.
 Location/Qualifiers

source
 1..1437
 /organism="Orygia pseudotsugata single capsid nuclear
 polyhedrosis virus"
 /db_xref="taxon:10450"
 1..377
 /gene="DNA polymerase"
 <1..377
 /gene="DNA polymerase"
 /codon_start=3
 /product="DNA polymerase"
 /protein_id="BA02565.1"
 /db_xref="GI:222201"
 /translation="AYPLRLEDAETMRISWKHTIINTPNELLEITGDEKDALAE
 CYSALIEKYMHOQADKKRALALVATIRKAPSASDASGRARAKGAPSDDESSSD
 EDAPCEPKCANNTFKCLYKAQ"
 405..409
 /note="late promoter"
 421..1386
 /gene="gp37"
 421..1386
 /gene="gp37"
 /codon_start=1
 /product="gp37 (spheroidin-like protein)"
 /protein_id="BA02566.1"
 /db_xref="GI:222202"
 /translation="MYKLCVAFALAVPVRPHGYLSTPYARQYKCFADGNEFWPDNG
 DGVDPEACRNAYKKFHRARYRAVGAPGGEAAQAQYMFQYAEYAAGVNRDELVK
 REVLPHILGCAANDRHALLFGKSGMDEPFNMRPDVLYNRYORASHFNVECTAV
 HEPSTFEVYTKFTMDRRSPVWNLELYTGNGSGIAPNGDAFCASGLVSIPIYVP
 YRPGFVAIVRWORIDPVGEGFTNCADIVFTENDEDRYRAAKAVRODLRQNCND
 CVEGPESCAPTRPQRANHLIRGAGHDQADGASVRETTDEL"

BASE COUNT 300 a 462 c 444 g 231 t
 ORIGIN

Query Match 72.4%; Score 21; DB 14; Length 1437;
 Best Local Similarity 82.8%; Pred. No. 4.6e+02;
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 caccgaagcgctgcgcacgacac 29
 |||||||||||||||||||||
 DB 259 CGCGCAAGCGCGCGCCAGCAGCAGC 287

RESULT 4
 LOCUS OP039145/c 5388 bp DNA linear VRL 18-JUL-1996
 DEFINITION Orygia pseudotsugata nuclear polyhedrosis virus DNA polymerase
 gene, complete cds, and ORF 66 gene, partial cds.
 ACCESSION U39145
 VERSION U39145
 KEYWORDS
 SOURCE Orygia pseudotsugata nuclear polyhedrosis virus.
 ORGANISM Orygia pseudotsugata single capsid nuclear polyhedrosis virus
 Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
 Nucleopolyhedrovirus.
 1 (bases 1 to 5388)
 Gross,C.H., Wolgamot,G.M., Russell,R.L., Pearson,M.N. and
 Rohmann,G.F.
 A 37-kilodalton glycoprotein from a baculovirus of Orygia
 pseudotsugata is localized to cytoplasmic inclusion bodies
 J. Virol. 67 (1), 469-475 (1993)
 93100831
 COMMENT 2 (bases 1 to 5388)
 Ahrens,C.H., Carlson,C. and Rohmann,G.F.
 Identification, sequence, and transcriptional analysis of lef-3, a
 gene essential for Orygia pseudotsugata baculovirus DNA replication
 Virology 210 (2), 372-382 (1995)
 95343549
 3 (bases 1 to 5388)
 Ahrens,C.H. and Rohmann,G.F.
 The DNA polymerase and helicase genes of a baculovirus of Orygia
 pseudotsugata

RAVRLNRAVADILKLAEDYADTAAYMAQDOPEASSRHHFATLNRMLLLIGVDDPAR
KRLNSVLARIEHLRVYNDABEVNVLSDGYEEKSKYISVQPEAQTNPASASQOQ
TSLPRQTSLEPROQASLPQOQPEQPEWSPSPFVHTTPTALPQTPPTPTDPSRP
SDEFLVYVGEKRAVPDTRKPPVPEHLKSRSSVATNAAGATPAVAPPPTPSAD
VTTSMPPPPPSADVTTSMPPPPVADLASMPPPPPPVADLASMPPPIINNA
INNLIDAMVAETNKAGNDSALLDOIKOGTKTKTOPADGAPATDPRSTLLSEIRQ
GKTLKIKKIDIOSSTOPLLKDVDTDTKTLINKFVYNIDRISKQEBEEDRLDTIT
KRPAVEHTDNGSTGNSDDMRD"

CDS
/note="ORF3: ph, AcMNPV ORF8 homolog"
/codon_start=1
/product="polynhedrin"
/protein_id="AAC59002.1"
/db_xref="GI:1911249"
/translation="MPTSYRPTTGRITTYVYDKYIKKNGSVYKNAKKRKHLEHEDE
KHLDPDHYMADEPFLGPKKQKLTLEKRNKVPDTMKLIVMNGSEYLETITR
VDSPEYVNDQVMDVFLVNNRPTPNPCYKFLAOLHLMDCDVPVPEVIRIPEST
VGMNNEYRISLAKGGGCPINNIHAETNSFVNRYIWNFYKPIYVYIGTDSSE
EILIEVLYEKYKEFPADAPLFTGPAY"
complement(3350..3541)
/note="ORF4"
/codon_start=1
/product="unknown"
/protein_id="AAC59003.1"
/db_xref="GI:1911250"
/translation="MRKNGASQYPEHLTTVHKRRLQLSCVWCDDQSRKSWPSAKGL
MEHMFCCRLITYOQKIKH"
complement(3698..4093)
/note="ORF5"
/codon_start=1
/product="unknown"
/protein_id="AAC59004.1"
/db_xref="GI:1911251"
/translation="MGASQNSECKITKLEFYRNSQYGAALDDEKDLHCLYDLEREVGA
HLNRADDAARRKCKAEKALKRVEITADPRHMLAEAAVACADDCGHWSTLSAQDDI
AREKEIYDRIRHQLKODSLKODRLKR"
complement(4125..4739)
/note="ORF6: left-2, AcMNPV ORF6 homolog"
/codon_start=1
/product="late expression factor 2"
/protein_id="AAC59005.1"
/db_xref="GI:1911252"
/translation="MERVWNPDAAGIDGLKRSETYLVDPHDFVGLTSLPYVENGLE
VRMSGRLIALIAPKROEPQAPRRFPQSRNNVCIKACADGASGLAVLARVSM
PMSKTMADISSAPRGMYRKREFFNCYLANVITCTKACIALGALLHFRYMDACY
GEVTHLIKADQYVYKSNCAKMKVYTKICQASMCGLNPICNF"
complement(4742..4975)
/note="ORF7: AcMNPV ORF5 homolog"
/codon_start=1
/product="unknown"
/protein_id="AAC59006.1"
/db_xref="GI:1911253"
/translation="MNRPTYRNTRAAYTDDYDRQOLRELNSLRSHHELCTRSATGFD
CNRPLEAGDRAPAYVAAANGGHSLICDKY"
5020..5460
/note="ORF8: AcMNPV ORF4 homolog"
/codon_start=1
/product="unknown"
/protein_id="AAC59007.1"
/db_xref="GI:1911254"
/translation="MPCARFVRSRLRTIOEFKEVYAVHVDHMLGRLLIDGKTSADV
RRFGFLSRNVLVSACMAVNOVYAPDATIMRHQPTTYFVNCOCCHAMADVPAPDDHS
IARYLAECAVAILDHPDVFGEETEGVELLEVORINAGDL"
complement(5669..6151)
/note="ORF9: ptp-2"
/codon_start=1
/product="protein tyrosine phosphatase 2"
/protein_id="AAC59008.1"
/db_xref="GI:1911255"
/translation="MYEASQIDEHVFVGYGDNEMALRIENHAIASVLSIDSDVA
PIRHAGLPVDHIIHYCEAPRTCAALPNNAPALDYVIRIIEGKRVILHVCYAGSR
SAAIAVYILMSROMPYODALNAVOSKRRAVINHFVPLATRCSTYFVNDLKPQIV

complement(6129..6791)
/note="ORF10: ptp-1, AcMNPV ORF1 homolog"
/codon_start=1
/product="protein tyrosine phosphatase 1"
/protein_id="AAC59009.1"
/db_xref="GI:1911256"
/translation="MLTSKRWAVITYRTSERNGCAKCTNGVRRRQVLPVSVARAM
FPRMHEYTAGCAVILEGTRLCRKYALEEYVTSDEDRWTASVLRHSLCAVI
DLNTIARYDGAQMVKGLYKIRVGRVAPDDDIATVEEIEVEFFRCPTMVAV
HHTHGLNRSGLVCRVWERLGVSPDIALAFETARGHKTERTNYLQDILLARKHVQ
PN"
6819..7814
/note="ORF11: AcMNPV ORF11 homolog"
/codon_start=1
/product="unknown"
/protein_id="AAC59010.1"
/db_xref="GI:1911257"
/translation="MSLSKLLVYAYGYNLPDHRGSGYHLRTVHEHLNTVYSN
ASCVRLIATVARGCLNSGHLCFDVAROLDYSEVAARLASMFRGCGATGICAMQFALA
DIDRHAPLARVGRKRANIETALDIADIPSDVTNNLGTIGREHHPFGSLARVADV
DPDIRADGMMYHKECVLTVMHLYACGAVPAGSVTRLDAVAKHIGFNDGNCAPATAV
VYGRFCAIGREHFAHKTACMHILFQMRNDLTPADERHPCFVIGFIRQCKDYTD
LRPHADALYIHGTTDRQNALFDLCCVNASDIDADCYDCVVKFYATONKKYKM"
complement(8180..9142)
/note="ORF12: AcMNPV ORF13 homolog"
/codon_start=1
/product="unknown"
/protein_id="AAC59011.1"
/db_xref="GI:1911258"
/translation="MFSMFGNMMAADEOVNAERDEOAYRYRVAVDORAHSLVRRDVE
RCHPFVRFKRVYDEDTAGRCSSVYDPCGKISHDLRNCNPROHVLNLEVLGAP
PAQPDLSGLSEFATKHLGLQLOQSFANKREVLAVANGANDHNDRLNLTETAVLHV
KTLNTNSDKFISAKRSKLEVGAREFQEDRLDTLKALNDCAPATPRAPVYPR
DYTKHPLAVFMGRVEDRGVQIAPFAGQDEHFEKRLLEEEDGADVORAPNPPLA
VHCJKEEPANGCHKIRRLPKKYLEIVDCAVNAVAKDIYKALINT"
complement(9094..9825)
/note="ORF13: left-1, AcMNPV ORF14 homolog"
/codon_start=1
/product="late expression factor 1"
/protein_id="AAC59012.1"
/db_xref="GI:1911259"
/translation="MAPCKYPERVQMMMDAIAVNSRRLAFMTDRPMVHAHNEFDS
AAOLFAYTVKNSISDVHYKPLEEGREVDADKDCDAKAEMLKYNVGAETMFEF
EGKEADVORINPSSNRGPHMLKFCGFKMDAPSLSEHNMVYKQAKLYSGDIRG
SFADVRAVAMHITGDAREDIYLRKWPVDVDRVCNNKQIRAPFSYNYGQV3RCL
TQQLQORIKGSAGCLAGTPTSK"
9893..11362
/note="ORF14: egt, AcMNPV ORF15 homolog"
/codon_start=1
/product="ectyosteroid UDP-glucosyltransferase"
/protein_id="AAC59013.1"
/db_xref="GI:1911260"
/translation="MVFLLIALTLTLAGRAASILAVLPYAISSHVYRAYVALYK
NCHNTYVIRKPDLDYAVODEGGRVQEDIDAMSQYKRLVASSGVFKRGVADETTV
TADNYGLILEMKDOFDNANVRFLSTRTFDVAVVEAFADYALFVGHLEPRAPVIGI
APGVGLAENERRRRAVRRPHLYPTFGAALLTRGGLSEWRLNEFELARRSDELL
KQGFCKSPITRIQLPDNYOILLNHPYVDNAPRPPVVOYLGGGLHQAOLPORDLA
PLERLNSVDGAVTVSFGSLDTNSHAFELQLOSSDDELADARPMPYCLPMKGOEFH
VALPRNVLAQKWFQTSOTAVLRKNVAFETQAGLOSSDDELADARPMPYCLPMKGOEFH
ARKLOQFGVARALDTAAVASAPOLQLAIEVLAIDGAVYARIDKLRAVVEHDAADDERA
VKEFERVIEFNDDVNMVPAKSLKTPTAANNAYSDFVFRPL"
11547..12140
/note="ORF15: AcMNPV ORF16 homolog"
/codon_start=1
/product="unknown"
/protein_id="AAC59014.1"
/db_xref="GI:1911261"
/translation="METADPPISTVAPPRKGAVCAYRTVTVTTSVSDSGNNEELTQ

Query Match 72.48; Score 21; DB 14; Length 131995;
Best Local Similarity 82.88; Pred. No. 2e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;


```

Oy      1  caccgaagcgcgcgcgcgcgcgcacac 29
        | | | | | | | | | | | | | | | |
Db      55017  cccgcaaaaggcgccgcccacgacgac 54989

RESULT  6
SC7H1/c 35654 bp  DNA  Linear  BCF 22-MAR-2000
LOCUS Streptomyces coelicolor cosmid 7H1.
DEFINITION A1021411
ACCESSION A1021411
VERSION 1
KEYWORDS abfB; aminotransferase; arabinofuranosidase; cbfM; cbfN; cbfO;
          cbfQ; chitinase; cobalt transport system; export protein;
          oxidoreductase; peptidylprolyl isomerase; sigma factor;
          transferase; xlnA; xylanase A.
SOURCE Streptomyces coelicolor A3(2).
ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
          Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 35654)
AUTHORS Redenbach,M., Kleser,H.M., Denapate,D., Eichner,A., Cullum,J.,
          Kinashi,H. and Hopwood,D.A.
          A set of ordered cosmids and a detailed genetic and physical map
          for the 8 mb Streptomyces coelicolor A3(2) chromosome
          Mol. Microbiol. 21 (1), 77-96 (1996)
TITLE 3 (bases 1 to 35654)
MEDIINE 97000351
REFERENCE 2 (bases 1 to 35654)
AUTHORS Murphy,L. and Harris,D.
          Unpublished
JOURNAL 3 (bases 1 to 35654)
AUTHORS Parkhill,J., Barrell,B.G. and Rajandream,M.A.
          Direct Submission
REFERENCE Submitted (20-JAN-1998) Streptomyces coelicolor sequencing project,
          Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
          CV10 1SA E-mail: barrellesanger.ac.uk Cosmids supplied by Prof.
          David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
          Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT Notes:
          Streptomyces coelicolor sequencing at The Sanger Centre is funded
          by the BBSRC.
          Details of S. coelicolor sequencing at the Sanger Centre are
          available on the World Wide Web.
          (URL: http://www.sanger.ac.uk/Projects/S.coelicolor/) CDS are
          numbered using the following system eg SC7B7.01c. SC (S.
          coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
          strand).
          The more significant matches with motifs in the PROSITE database
          are also included but some of these may be fortuitous. The length
          in codons is given for each CDS.
          Usually the highest scoring match found by fasta -o is given for
          CDS which show significant similarity to other CDS in the database.
          The position of possible ribosome binding site sequences are given
          where these have been used to deduce the initiation codon. Gene
          prediction is based on positional base preference in codons using a
          specially developed Hidden Markov Model (Krogh et al., Nucleic
          Acids Research, 22(22):4768-4778(1994)) and the Frameplot program
          of Bibb et al., Gene 30:157-66(1994) as implemented at
          http://www.nih.gov.jp/
          jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the
          correct initiation codon. Where possible we choose an initiation
          codon (atg, gtg, ttg or att) which is preceded by an upstream
          ribosome binding site sequence (optimally 5-13bp before the
          initiation codon). If this cannot be identified we choose the most
          upstream initiation codon.
          IMPORTANT: This sequence MAY NOT be the entire insert of the
          sequenced clone. It may be shorter because we only sequence
          overlapping sections once, or longer, because we arrange for a
          small overlap between neighbouring submissions. Cosmid 7H1 lies
          between 10k5 and 7B7 on the AseI-B genomic restriction fragment.
          Location/Qualifiers
            1..35654
              /organism="Streptomyces coelicolor A3(2)"
              /strain="A3(2)"

```

misc_feature

CDS

misc_feature

misc_feature

RBS

CDS

gene

gene

gene

gene

gene

gene

gene

gene

gene

Estimated insert size: 122355; sum-of-contrigs estimation
Quality coverage: 0x in Q20 bases; agarose-ff estimation
Quality coverage: 2.1x in Q20 bases; sum-of-contrigs estimation

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
consists of 74 contrigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contrigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 5209: contrig of 5209 bp in length
5210 5309: gap of unknown length
5310 10338: contrig of 5029 bp in length
10339 10438: gap of unknown length
10439 14826: contrig of 4388 bp in length
14827 14926: gap of unknown length
14927 19138: contrig of 4212 bp in length
19139 19238: gap of unknown length
19239 23107: contrig of 3869 bp in length
23108 23207: gap of unknown length
23208 26969: contrig of 3762 bp in length
26970 27070: gap of unknown length
27071 32006: contrig of 4937 bp in length
32007 32106: gap of unknown length
32107 34680: contrig of 2574 bp in length
34681 34780: gap of unknown length
34781 37409: contrig of 2629 bp in length
37410 37509: gap of unknown length
37510 41235: contrig of 3726 bp in length
41236 41336: gap of unknown length
41337 44424: contrig of 3089 bp in length
44425 44524: gap of unknown length
44525 48770: contrig of 4246 bp in length
48771 48870: gap of unknown length
48871 52031: contrig of 3161 bp in length
52032 52131: gap of unknown length
52132 55038: contrig of 2907 bp in length
55039 55138: gap of unknown length
55139 58265: contrig of 3127 bp in length
58266 58365: gap of unknown length
58366 61071: contrig of 2706 bp in length
61072 61171: gap of unknown length
61172 62802: contrig of 1631 bp in length
62803 62902: gap of unknown length
62903 65216: contrig of 2314 bp in length
65217 65316: gap of unknown length
65317 68673: contrig of 3357 bp in length
68674 68773: gap of unknown length
68774 71777: contrig of 3004 bp in length
71778 71877: gap of unknown length
71879 75040: contrig of 3163 bp in length
75041 75140: gap of unknown length
75141 77111: contrig of 1971 bp in length
77112 78277: contrig of 1716 bp in length
78278 79027: gap of unknown length
79028 81095: contrig of 2068 bp in length
81096 81195: gap of unknown length
81196 83729: contrig of 2534 bp in length
83730 83829: gap of unknown length
83830 85747: contrig of 1918 bp in length
85748 85847: gap of unknown length
85849 88908: contrig of 3061 bp in length
88909 90558: gap of unknown length
90559 90658: contrig of 1550 bp in length
90659 93270: contrig of 2612 bp in length
93271 93370: gap of unknown length
93371 95096: contrig of 1726 bp in length

95097 95197 95196: gap of unknown length
95197 96641: contrig of 1445 bp in length
96642 96741: gap of unknown length
96742 98704: contrig of 1963 bp in length
98705 98804: gap of unknown length
98805 100547: contrig of 1742 bp in length
100547 100646: gap of unknown length
100647 101882: contrig of 1236 bp in length
101883 101982: gap of unknown length
101983 103008: contrig of 1026 bp in length
103009 103108: gap of unknown length
103109 104820: contrig of 1712 bp in length
104821 104921: gap of unknown length
104921 106395: contrig of 1475 bp in length
106396 106496: gap of unknown length
106496 108468: contrig of 1973 bp in length
108469 108568: gap of unknown length
108569 110201: contrig of 1633 bp in length
110202 110301: gap of unknown length
110302 111838: contrig of 1537 bp in length
111839 111939: gap of unknown length
111939 113656: contrig of 1718 bp in length
113657 113756: gap of unknown length
113757 115783: contrig of 2027 bp in length
115784 115883: gap of unknown length
115884 117732: contrig of 1849 bp in length
117733 117832: gap of unknown length
117833 119791: contrig of 1959 bp in length
119792 119891: gap of unknown length
119892 121704: contrig of 1813 bp in length
121705 121804: gap of unknown length
121805 123619: contrig of 1815 bp in length
123620 123719: gap of unknown length
123720 125324: contrig of 1605 bp in length
125325 125424: gap of unknown length
125425 127220: contrig of 1796 bp in length
127221 127320: gap of unknown length
127321 128937: contrig of 1617 bp in length
128938 129037: gap of unknown length
129038 130061: contrig of 1024 bp in length
130062 130161: gap of unknown length
130162 131556: contrig of 1395 bp in length
131557 131656: gap of unknown length
131657 132821: contrig of 1165 bp in length
132822 132921: gap of unknown length
132922 134359: contrig of 1438 bp in length
134360 134459: gap of unknown length
134460 136156: contrig of 1697 bp in length
136157 136256: gap of unknown length
136257 137579: contrig of 1323 bp in length
137580 137679: gap of unknown length
137680 138787: contrig of 1108 bp in length
138788 138887: gap of unknown length
138889 140046: contrig of 1159 bp in length

Query Match 69.0%; Score 20; DB 2; Length 162623;
Best Local Similarity 82.1%; Pred. No. 4; Se+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

07 2 acgcgaagcgctcgccagcaagac 29
db 110089 acgcccggccgctcgccagcaagac 110062

RESULT 8
BC011138/c BC011138 2053 bp mRNA linear PRI 30-JUL-2001
LOCUS Homo sapiens, clone MGC:17456 IMAGE:3448753, mRNA, complete cds.
DEFINITION
ACCESSION BC011138
VERSION BC011138.1 GI:15029829
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens

repeat_region /rpt_family="AluSc"
745..911
/rpt_family="Alu"
repeat_region 1803..1993
/rpt_family="MER20"
complement(join(<2487..2603,2694..2856,3741..3886,
4321..4416,7302..>7579))
/note="Hypothetical partial protein; Most similar to
YB3-CSDPO HYPOTHETICAL 60.9 KD PROTEIN C2F12.12C IN
CHROMOSOME II gnl|PID325370 (297211) hypothetical
protein [Schizosaccharomyces pombe]; 5'-end of
hypothetical gene not discernible in genomic sequence at
this time"
/codon_start=3
/evidence=not-experimental
/product="R31449_3"
/protein_id="AAC32903.1"
/db_xref="GI:3461802"
/translation="RSSSPRAAALSSQOSLQERLRERKQOEIMKAFETPEEK
BARLAKREAKRRKREKMGEEYMGTTNDNPGDNMLGTFIWNKALKKGISHL
EEKELKRNKRIOEDNRLQKQKQLRERERKAMRNQLEMLQREKAEKFTWEE
QEDNFHQAKIRSKIRIDGRAPIDILAKYISAEDDLAEHMEHEPTFNGILTVAD
MEDLEIOVYMEEGOGKNADFEWDMTTTFEDELISKRLKLEASGKPG"
misc_feature complement(2489..2603)
/note="Predicted exon, program: graal2exons_human_1.3,
frame: 2, quality: excellent, score: 100.000"
misc_feature complement(2694..2856)
/note="Predicted exon, program: graal2exons_human_1.3,
frame: 2, quality: excellent, score: 100.000"
misc_feature complement(3741..3886)
/note="Predicted exon, program: graal2exons_human_1.3,
frame: 0, quality: excellent, score: 100.000"
misc_feature complement(4321..4416)
/note="Predicted exon, program: graal2exons_human_1.3,
frame: 1, quality: excellent, score: 99.000"
repeat_region complement(5352..5489)
/rpt_family="LINE2"
5698..5997
/rpt_family="AluSx"
repeat_region 6002..6360
/rpt_family="MLTIC"
repeat_region 6568..6858
/rpt_family="AluDb"
repeat_region 6862..7152
/rpt_family="AluY"
complement(7302..7579)
/note="Predicted exon, program: graal2exons_human_1.3,
frame: 2, quality: excellent, score: 96.000"
repeat_region complement(7478..7538)
/rpt_family="(CAG)n"
repeat_region complement(7819..8275)
/rpt_family="LINE2"
repeat_region complement(8516..8805)
/rpt_family="AluIo"
repeat_region 9063..9360
/rpt_family="AluSx"
repeat_region 9611..9994
/rpt_family="LINE2"
repeat_region complement(10187..10279)
/rpt_family="(CGG)n"
repeat_region complement(11027..11073)
/rpt_family="MIR"
repeat_region 11376..11675
/rpt_family="AluSx"
repeat_region 11697..11989
/rpt_family="AluSg"
repeat_region 12261..12559
/rpt_family="AluSx"
repeat_region complement(12660..12847)
/rpt_family="AluSg/x"
repeat_region 13102..13401
/rpt_family="AluSx"
repeat_region 13406..13703

misc_feature /rpt_family="AluY"
13795..15474
/note="DSD similarity to overlapping ESTs:
(13795..14230) AA621023 ag3c09.s1 Soares testis NHT Homo
sapiens cDNA clone 1056208 3'; Score: 856 Identity:
432/436 (99%).-(13795..14267) A1018772 ov32e07.x1
Soares testis NHT Homo sapiens cDNA clone IMAGE:1639044 3'
similar to contains TARL13 P785 repetitive element ;
Score: 938 Identity: 471/473 (99%).-(14560..14100)
AA128938 z011b12.t1 StrataGene neuroepithelium NT2RAMT
937234 Homo sapiens cDNA clone 567359 5'; Score: 886
Identity: 453/459 (98%).-(14913..14564) W94459 zei2e03.r1
Soares fetal heart NBH19W Homo sapiens cDNA clone 358780
5'. Score: 668 Identity: 345/350 (98%).-(15439..14978)
R35135 y960e12.r1 Homo sapiens cDNA clone 36877 5'. Score:
766 Identity: 447/476 (93%).-(15474..15189) AA323207
EST25973 Cerebellum II Homo sapiens cDNA 5' end; Score:
530 Identity: 278/284 (97%).-Additional EST matches:
AA975069, AA128939, AA524185, AA483347, H67626, H73320,
M78980, and many others. . . ."
complement(13798..>37203)
/note="Encodes human homolog of mouse
phosphatidylinositol-4-phosphate 5-kinase I gamma"
/product="Homo sapiens mRNA for KIAA0589 protein"
complement(15645..16081)
/note="DSD similarity to overlapping ESTs:
(15842..15645) 242707HSC0G021.H sapiens partial cDNA
sequence; clone c-0rg02. Score: 388 Identity: 196/198
(98%).-(15884..15653) T31346 EST34513 Homo sapiens cDNA 5'
end similar to None; Score: 448 Identity: 228/232
(98%).-(16081..15693) AA446775 zw89d06.r1 Soares total
fetuss NB2Hr8 9W Homo sapiens cDNA clone 784139 5'; Score:
720 Identity: 377/387 (97%).~"
misc_feature 16162..16753
/note="DSD similarity to AA527155 n107c01.s1 NCI-CGAP_Br2
Homo sapiens cDNA clone IMAGE:967296; Score: 958
Identity: 546/592 (92%).~"
complement(join(16781..16783,17051..17134,22498..22630,
25319..25423,26521..26553,26857..26995,27701..27865,
29588..29672,30952..31000,32239..32322,35440..35645,
36904..>37203))
/function="putative lipid kinase"
/note="Human homolog of mouse phosphatidylinositol-
4-phosphate 5-kinase I gamma"
/codon_start=1
/product="KIAA0589, partial CDS"
/protein_id="AAC32904.1"
/db_xref="GI:3461803"
/translation="NLNQNPRLLPKFGLCYVSGGKNIKVYVNNNLLPVRVKNHLK
FDLGSYKRRASKEKESPFYKLDPMQDMEGLLDPTFSALVKTORCLYL
ESFKIMYSILIGVHNIDQHREROAGAOSTSEKRRYQKALYSTRAMESTOGGAAR
GEATESDDTMGIPRAVNGRGRLLHLHGTIDILDSYRTRKLEHTWALVHDGTVSV
HRPSYIAERPFKSNVFRKNSSLSKSPSGGALAVKPGPTAAASASO1PSE
REEAOYDLRGARSYPTLEDEGRPDLCPTSPSEFATASLATLSSISIPRPS
ETSEQPRYRRRTQSGODGRPEEPEDQCTTVQVEPCSEIVVPRKEDGVS
SPAGASAAVEYETASQASDEGAPASQASDEBAPADTDIYPTPERSGWSVPLHYSNO
APASGDSGSDP"
complement(18117..18182)
/note="Predicted exon, program: graal2exons_human_1.3,
frame: 2, quality: good, score: 52.000"
misc_feature complement(18733..18853)
/note="Predicted exon, program: graal2exons_human_1.3,
frame: 0, quality: good, score: 68.000"
repeat_region 18879..19014
/rpt_family="AluSx"
repeat_region 19016..19311
/rpt_family="AluSx"
repeat_region 19312..19613
/rpt_family="AluY"
repeat_region 19618..19790
/rpt_family="AluSx"
repeat_region 20316..20356

misc_feature complement(21150..21227)

Query Match 68.3%; Score 19.8; DB 9; Length 38209;
 Best Local Similarity 91.3%; Pred. No. 7.1e+02;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 gcaagcgctgcgcgcgcac 26
 |||||||||||||||
 Db 14718 GCAAGCGCTGC GC GC GCAC 14740

RESULT 13
 AC068475 175625 bp DNA linear HTG 01-JUN-2000
 LOCUS Homo sapiens chromosome 8 clone RP11-383B15 map 8, WORKING DRAFT
 DEFINITION SEQUENCE, 38 unordered pieces.
 AC068475
 AC068475.2 GI:8140670
 HTG: HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE
 1 (bases 1 to 175625)
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 Homo sapiens chromosome 8, clone RP11-383B15
 Unpublished
 2 (bases 1 to 175625)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
 Bonoslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,
 Campagnolo, A., Castle, A., Chepel, Y., Colangelo, M., Collins, S.,
 Collamore, A., Cooke, P., Dearlano, K., Dewar, K., Diaz, J. S.,
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Glade, S., Goyette, M., Graham, L.,
 Grand-pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,
 Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., Larocque, K., Lamazares, R., Landers, T., Lehoczy, J.,
 Levine, R., Lieu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N.,
 McCarthy, M., McKean, P., McGuck, A., McKernan, K., McPheters, R.,
 Melidiri, J., Meneus, L., Milnova, T., Miranda, C., Mlenga, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,
 Plesni, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Testa, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, D.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE
 JOURNAL
 COMMENT

Submitted (02-MAY-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jun 1, 2000 this sequence version replaced gi:7677958.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIRB
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center project name: 19767
 Center clone name: 383.B.15

Summary Statistics
 Sequencing vector: M13; M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 15681 bases at least Q40
 Consensus quality: 166139 bases at least Q30
 Consensus quality: 169716 bases at least Q20
 Insert size: 187000; agarose-1p

Insert size: 171925; sum-of-contigs
 Quality coverage: 3.7 in Q20 bases; agarose-1p
 Quality coverage: 4.0 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
 consists of 38 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1 1032: contig of 1032 bp in length
 1033 1132: gap of 100 bp
 1133 2147: contig of 1015 bp in length
 2148 2247: gap of 100 bp
 2248 3601: contig of 1354 bp in length
 3602 3701: gap of 100 bp
 3702 5049: contig of 1348 bp in length
 5050 5149: gap of 100 bp
 5150 6694: contig of 1545 bp in length
 6695 6794: gap of 100 bp
 6795 7668: contig of 1074 bp in length
 7669 7968: gap of 100 bp
 7969 9659: contig of 1691 bp in length
 9660 9759: gap of 100 bp
 9760 11604: contig of 1845 bp in length
 11605 11704: gap of 100 bp
 11705 13073: contig of 1369 bp in length
 13074 13173: gap of 100 bp
 13174 15155: contig of 1962 bp in length
 15156 15255: gap of 100 bp
 15256 17626: contig of 2371 bp in length
 17627 17726: gap of 100 bp
 17727 20315: contig of 2589 bp in length
 20316 20415: gap of 100 bp
 20416 21588: contig of 1173 bp in length
 21589 21688: gap of 100 bp
 21689 23898: contig of 2210 bp in length
 23899 23998: gap of 100 bp
 23999 26021: contig of 2023 bp in length
 26022 26121: gap of 100 bp
 26122 27520: contig of 1399 bp in length
 27521 27620: gap of 100 bp
 27621 29883: contig of 2263 bp in length
 29884 29983: gap of 100 bp
 29984 33054: contig of 3071 bp in length
 33055 33154: gap of 100 bp
 33155 36994: contig of 3840 bp in length
 36995 37094: gap of 100 bp
 37095 40456: contig of 3362 bp in length
 40457 40556: gap of 100 bp
 40557 44872: contig of 4316 bp in length
 44873 44972: gap of 100 bp
 44973 49433: contig of 4461 bp in length
 49434 49533: gap of 100 bp
 49534 54009: contig of 4476 bp in length
 54010 54109: gap of 100 bp
 54110 57886: contig of 3777 bp in length
 57887 57986: gap of 100 bp
 57987 63137: contig of 5151 bp in length
 63138 63237: gap of 100 bp
 63238 68261: contig of 5024 bp in length
 68262 68361: gap of 100 bp
 68362 73156: contig of 4795 bp in length
 73157 73256: gap of 100 bp
 73257 79722: contig of 6466 bp in length
 79723 79822: gap of 100 bp
 79823 86499: contig of 6677 bp in length
 86500 86599: gap of 100 bp
 86600 91151: contig of 4552 bp in length
 91152 91251: gap of 100 bp
 91252 98153: contig of 6902 bp in length

* 98154 98253: gap of 100 bp
* 98254 103322: contig of 4969 bp in length
* 103223 103322: gap of 100 bp
* 103323 110113: contig of 6791 bp in length
* 110114 110213: gap of 100 bp
* 110214 116818: contig of 6605 bp in length
* 116819 116918: gap of 100 bp
* 116919 127763: contig of 10845 bp in length
* 127764 127863: gap of 100 bp
* 127864 137894: contig of 10031 bp in length
* 137895 137994: gap of 100 bp
* 137995 149790: contig of 11796 bp in length
* 149791 149890: gap of 100 bp
* 149891 175625: contig of 25735 bp in length.
Location/Qualifiers
1. 175625
/organism="Homo Sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-383B15"
/clone_11b="RPC1-11 Human Male BAC"
1. 1032
/note="assembly-fragment"
1133. 2147
/note="assembly-fragment"
2248. 3601
/note="assembly-fragment"
3702. 5049
/note="assembly-fragment"
5150. 6694
/note="assembly-fragment"
6795. 7868
/note="assembly-fragment"
7969. 9659
/note="assembly-fragment"
9760. 11604
/note="assembly-fragment"
11705. 13073
/note="assembly-fragment"
13174. 15155
/note="assembly-fragment"
15256. 17626
/note="assembly-fragment"
17727. 20315
/note="assembly-fragment"
20416. 21588
/note="assembly-fragment"
clone_end:T7
vector_side:right
21689. 23898
/note="assembly-fragment"
23999. 26021
/note="assembly-fragment"
26122. 27520
/note="assembly-fragment"
clone_end:SP6
vector_side:right
27621. 29883
/note="assembly-fragment"
29884. 33054
/note="assembly-fragment"
33155. 36994
/note="assembly-fragment"
37095. 40456
/note="assembly-fragment"
40557. 44872
/note="assembly-fragment"

Query Match 68.3%; Score 19.8; DB 2; Length 175625;
Best Local Similarity 91.3%; Pred. No. 5.3e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 gcaagcgcgctgcgcacacac 26
|||||
Db 118633 GCAAGCGCGCTGCACAC 118655

RESULT 14
AE004513/C
LOCUS
DEFINITION
Pseudomonas aeruginosa PA01, section 74 of 529 of the complete
genome.
ACCESSION
AE004513 AE004091
VERSION
AE004513.1 GI:9946669
KEYWORDS
SOURCE
Pseudomonas aeruginosa.
Pseudomonas aeruginosa.
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
1 (bases 1 to 15470)
Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,
Hickey,M.J., Brinkman,F.S.L., Huftagle,W.O., Kowalik,D.J., Lagrou,M.,
Garber,R.L., Goltzy,L., Tolentino,E., Westbrock-Wadman,S., Yuan,Y.,
Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,
Smith,K., Spencer,D., Wong,G.K., Wu,Z., and Paulsen,I.T.
Complete genome sequence of Pseudomonas aeruginosa PA01, an
opportunistic pathogen
Nature 406 (6799), 959-964 (2000)
20437337
2 (bases 1 to 15470)
Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,
Hickey,M.J., Brinkman,F.S.L., Huftagle,W.O., Kowalik,D.J.,
Lagrou,M., Garber,R.L., Goltzy,L., Tolentino,E.,
Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Sater,M.H.,
Hancock,R.E.W., Lory,S. and Olson,M.V.
Direct Submission
Submitted (16-MAY-2000) Department of Medicine and Genetics,
University of Washington Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA

TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
Pseudomonas aeruginosa
PA01, section 74 of 529 of the complete
genome.
AE004513 AE004091
VERSION
AE004513.1 GI:9946669
KEYWORDS
SOURCE
Pseudomonas aeruginosa.
Pseudomonas aeruginosa.
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
1 (bases 1 to 15470)
Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,
Hickey,M.J., Brinkman,F.S.L., Huftagle,W.O., Kowalik,D.J.,
Lagrou,M., Garber,R.L., Goltzy,L., Tolentino,E.,
Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Sater,M.H.,
Hancock,R.E.W., Lory,S. and Olson,M.V.
Direct Submission
Submitted (16-MAY-2000) Department of Medicine and Genetics,
University of Washington Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA

FEATURES
Source

1. 15470
/organism="Pseudomonas aeruginosa"
/strain="PA01"
/db_xref="taxon:287"
126. 878
/gene="PA0780"
126. 878
/gene="PA0780"
/codon_start=1
/transl_table=1
/product="probable transcriptional regulator"
/protein_id="AAG04169.1"
/db_xref="GI:9946670"

126. 878
/gene="PA0780"
126. 878
/gene="PA0780"
/codon_start=1
/transl_table=1
/product="probable transcriptional regulator"
/protein_id="AAG04169.1"
/db_xref="GI:9946670"
/translation="MFDRLRLRPDSQHTAHENHQLVMSLAGRAFEVNGGGEVCR
MRACIVPEAGHVFAGVGNMMLIIDGENDPPEDDPLARLEAPRYPLADPQN
LISVAGAEIARVGSPLTRALGVILRALHILRFGEOPPLAASGPDLDELADHAYR
NIARITVAELAOVACLSPSHFHOFDSGCTPHOYLKVRDCAARLLRESNPLV
RIAEPCFSOSATITAMRYLGLTPSLKRG"
complement(999. 3062)
/gene="PA0781"
complement(999. 3062)
/gene="PA0781"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAG04170.1"
/db_xref="GI:9946671"

126. 878
/gene="PA0780"
126. 878
/gene="PA0780"
/codon_start=1
/transl_table=1
/product="probable transcriptional regulator"
/protein_id="AAG04169.1"
/db_xref="GI:9946670"
/translation="MFDRLRLRPDSQHTAHENHQLVMSLAGRAFEVNGGGEVCR
MRACIVPEAGHVFAGVGNMMLIIDGENDPPEDDPLARLEAPRYPLADPQN
LISVAGAEIARVGSPLTRALGVILRALHILRFGEOPPLAASGPDLDELADHAYR
NIARITVAELAOVACLSPSHFHOFDSGCTPHOYLKVRDCAARLLRESNPLV
RIAEPCFSOSATITAMRYLGLTPSLKRG"
complement(999. 3062)
/gene="PA0781"
complement(999. 3062)
/gene="PA0781"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAG04170.1"
/db_xref="GI:9946671"
/translation="MSSSGGLFSPRPLPLPLALACLIVSGETIGADGRSELPQSVI
TANPLGNESPPATPSVLEDELTRKSGSETLNGIPGVSTYFGGASRPVIRGMD
GDRILRLNGVGLDASLSYDAHPDPSVRLTYRGPALILYGNALIGVYNSF
DNITPSRPVGHISGELRKGADITRSKGALEPAQGNFALHVDASREFNVRIPG
YAHSRORIDSDGKRVHVSNGDGGVAGVSYNHEHGVAGLSYSGDSNYSNPAAE
DVALRKMQDRIYAFASEIRDLSEPFSLKIDAAVTKYEHKEIEDGETGTFKMEGYEG

```

gene
3526..6708
/ gene="puta"
/ note="PA0782"
3526..6708
/ gene="puta"
/ codon_start=1
/ transl_table=11
/ product="proline dehydrogenase Puta"
/ protein_id="AAG04171.1"
/ db_xref="GI:9946672"
/ translation="MFKASHVLQGEDQTRSNAEFPPVYSANYVADEAYIGELLQALD
PGDGEIERNRARALIETVRSRNDVDTLDALLRQYSLDTQEGIMLCIALLRYP
DAATADALIRKLTSAADQRLGSDNVLFNFAAGLVMTGKVNIDERTGRAPSVY
SLVQSGSEPVRAAMQAMTKKOPVIGRTISEALKNGRCRCOGTYTSPDMGEA
ALTADAEKYMDYRQADITVGAEPYGGPPRSISIKLSAHPRYEPAQREYVLSL
FANVLEAVARKLVNGITIDAEADRLSELLEYKLRDPAIANGWEEELVYQAS
KROLPIVWVLTIGRELGARIPRLVKGAWDEIKOCOVQAGDYPYTRKECTDTS
YLACARVLTSEHTRGVYPOFASHNATVTAILADEARAGAGEDEFEFORLHGNG
DALYDVIKRYRNRVYAPVGAHKDLPLYLRLLNGANSSFYHKLVDRVPEVL
IQRPVQLQKFTLANDRIPLPAIIPGAGKNSQGINNIONMMELELAKKPLHTQ
WQAPVYNGSKLSGPAHEKRCPTDLERLVSVYASAEQAKALDVLOAAMPBSAP
VDERAGILERLADLLEAQRGELMALCYLLEKSLQDIDEIREADVCRITYAQAQR
LYREBELRGPERNELFEGRGIFACVSPNNPPLAIFLQIISALVAGNVTLOAPAQ
TSILAAVTEIMFEAGLPKDYIALPDGATLGVCRDARVAGVAFSGRSTDRINII
ROLAEKGATLITLAEFGONAMVDSFALPEQYIKDAVOSAFPSAGORCSALVMYV
OQDAERVTEILKGAAELKYGPTSEVSDVPYIDAEKAGLEQHIALLAKKLLA
ETKYPGKLNHRYAPAPVAFISGIDELKKNNGVILHVRYAAGLDEKVOAINTGSG
LTMGVSHRNEETARKIEELARVGNITYNRQIGAVYGVYPGGHLSSTGKACAGPVY
LIRFVSERTSVTFAVGNSLSTLADAE
7020..8540
/ gene="putp"
/ note="PA0783"
7020..8540
/ gene="putp"
/ codon_start=1
/ transl_table=11
/ product="sodium/proline symporter Putp"
/ protein_id="AAG04172.1"
/ db_xref="GI:9946673"
/ translation="MSVNTPTLIPFVITYIAMVILGIAVRSNNNSDYILGSRSLSS
FVTLASGASDMGMLMGLPAVYLSLSLSWYALIGLYGATLNMPLPAGRIYQRE
HNGNALILPYFTNREDNSRLRLTESALVILVFETTCASGIVAGARLPSTEGLSY
ETALMAGAAATIAVTFIGFLAVSWTDVQASIMIFLITLPIYVLAATGVEPFTA
IELADATSPMLKGSFIVISLMAWGLYFGPHILARMAADVSKSPARIRISM
IMWICLGAVAVGFGIAFGAOPAGVSENERVEIELAKILENPIAGVLSLAI
LAAVMTLSGOLLVCSALTEDFYAKRGAOSOLELWVGAMVLLVAVIAVTLAS
PENRVGLVSTANAGGAFGPLYVPSLIMKRMTRNKAALAGMVGAAVYIMKNTIG
TGLYIIPGLFASVAIVFSLGKAPSTMLKRFDAEDETREAHQSGAAVYOR"
complement(8685..9614)
/ gene="PA0784"
complement(8685..9614)
/ gene="PA0784"
/ codon_start=1
/ transl_table=11
/ product="probable transcriptional regulator"
/ protein_id="AAG04173.1"
/ db_xref="GI:9946674"
/ translation="MEIPWASIPLTALRTPESAROSFRKAAEELAVTPAVSHOI
KSLKAWGLVPLFORLPKARKLIDCCGRIPRSIHALLEISQSVDTLRQRSYNTVS
TTPAALMLWVPLRGRIYAAHPQINRLDTVC EYVDLHQAISIDVIRSLDDHPRNY
GLCLDECFAYGSPAVYLAARQPTLISVMHNSRLYAHGEAWCAOAGDWMGEG
PVIRQYDEEHAILOAAIAGGIVLASIIVSVSAGILRPYADISVAGYSAICY
PGRERHPVKAFFEMLRLEARQSGHACRLERDA"
9741..10379
/ gene="PA0785"
9741..10379

```

```

/ gene="PA0785"
/ codon_start=1
/ transl_table=11
/ product="probable acyl carrier protein phosphodiesterase"
/ protein_id="AAG04174.1"
/ db_xref="GI:9946675"
/ translation="MSRIILAVHASPGRGSRRLAEVLAAYREAHQAVAREYG
RVPLPATEAFVAAAFHPQPEQRSLAMQADALASDQVGELEPDSULVITPMNTSV
PSGLKANTIDQIVRLGTFEDVLDNGVAVQRPRLGRKALIVTSKRGHGFSGNQAM
NHADPWLRKALFGIDETVVAABGERSGSRSPEDSCDEARQLALARSA"
10389..10694
/ gene="PA0786"
10389..10694
/ gene="PA0786"
/ codon_start=1
/ transl_table=11
/ product="probable transporter"
/ protein_id="AAG04175.1"
/ db_xref="GI:9946676"
/ translation="MAFHLVLAAPAEVAFPMGKKSNGFGRIMPSLTVVAATGTY
FLTALRELVSVAIPYIWTALISLTVEFLGVLLGSESLAVKLVSVGLTVAGVGLK"
10750..11913
/ gene="PA0787"
10750..11913
/ gene="PA0787"
/ codon_start=1
/ transl_table=11
/ product="hypothetical protein"
/ protein_id="AAG04176.1"
/ db_xref="GI:9946677"
/ translation="MLTTTAAVSYRTLDVYMPRLNLTGANGAKSNVYRALRL
ADSAAGEVPSLAEGSLPSTLWAGEDISRAMKSGQAPLEGSVRQSNVLELPGAGE
DFGYCIDLGEPESQSKFGIDVPIKRECIMAGPLRANVLRORPLRAGGGRB
EVAQHLATEPDSMIGAIVPQVPELIGIRIRISMRYPHFPSDAEAPARPOIGTH
PVLGADADALAAALQITREIGDAALDEAVDAFPESRLIDNPGLGREVILEQPL
LBRRTALSPGTSLRYLITLALISPRPLADVINEPTEIHLDDLPAIGRIVGQAO
HSQVTVSHARIVATLEHRECHSGLEKDEGEMRIQGLREIDRRAMVYVR"
complement(11918..15043)
/ gene="PA0788"
complement(11918..15043)
/ gene="PA0788"
/ codon_start=1
/ transl_table=11
/ product="hypothetical protein"
/ protein_id="AAG04177.1"
/ db_xref="GI:9946678"
/ translation="MNCATAGASMGASERTNDPSELQPCSLPASHSPKSRRLQRL
MLIGALPILAAVGIIVVILHEAFTSRLOARELRYAATLDYEVROGSEALILPPADGP
DRLRGYOLIPSEMRDLYDRDYAITROAHFSPALMYAHRLEFPYAEKAOGLIADG
KQVPIYDRFYQRRRYANPASTPLVQSLIFTEINDLIDHERPIANAVDMGRTQAA
LSQVGMKIGFSAHSSGSGTATQTEKTRBSAGRTSGTGKLRQMTSASVRSYRSGPA
NPARQODILVLYLNSVPLSAPGYEGVTLADGLMVTGADYRQVGBALDGRAGLAAQ
GIALROVALIAHRRSPFLAERGBDELDMTDSHLVLTQAGVTEAPLRDALAOK
LAFDRPSQPTVOPLPNPKVTLARTLARGMLGVLYLDLRLDRLANTYLLQHELQSV
GVALROVALIAHRRSPFLAERGBDELDMTDSHLVLTQAGVTEAPLRDALAOK
TAYLQKLDAPYAAOGLIGERLITPTSRVSVYSEFTLEERTPGNVOYVQDNTDOP
PDINGSKTEIGSTAKRVLVASYLVEVVAOIHDDYGMGVAELRKVEVEPDLFTLMGI
DIVASRDRDLISAMLOAMERKYSAPYSPPTGGCLHTFNNPKEDNGRPMLEAL
RSLINLPEVRLRLDLIRHDIYQNAKSRVOLADDDPPRAYLDRTYDKSSQYVLEAF
WVYTRDKANOARELFLDGLRPMVPRVLAAILHYLQPPQADLASFSATFLREPLVSGSLTD
KRAELYERYGKFNLDQGYARVAPRLMILGYLOKQOATFGVAAGSAEKE
VYGMLEKSRHKNARDKRIRLILEVEAFLLDHQWKKLGYPPDHLVPSYATVAGSSGDR

```

```

Query Match 67.6%; Score 19.6; DB 1; Length 15470;
Best Local Similarity 84.6%; Pred. No. 1e+03;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 cagcgaagggcgctggcgaagccac 26
||||| ||| ||||| |||
Db 3826 CAGGCAACAGCGCTGGCCAGACAC 3801

```

```

RESULT 15
AE008794

```

LOCUS AE008794 26591 bp DNA linear BCT 25-OCT-2001
DEFINITION *Salmonella typhimurium* LT2, section 102 of 224 of the complete genome.
ACCESSION AE008794 AE006468
VERSION AE008794.1 GI:16420649
KEYWORDS
SOURCE
ORGANISM *Salmonella typhimurium* LT2.
Salmonella typhimurium LT2.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; *Salmonella*.
REFERENCE
AUTHORS 1 (bases 1 to 26591)
McLelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W.,
Latreille, P., Courtney, L., Portolick, S., All, J., Dante, M., Du, F.,
Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A.,
Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Florea, L., Miller, W.,
Stonking, T., Nhan, M., Waterston, R. and Wilson, R.K.
TITLE Complete genome sequence of *Salmonella enterica* serovar *Typhimurium* LT2
JOURNAL Nature 413 (6858), 852-856 (2001)
PUBMED 11677609
2 (bases 1 to 26591)
The *Salmonella typhimurium* Genome Sequencing Project.
AUTHORS Direct Submission
TITLE Submitted (29-MAR-2001) Genome Sequencing Center, Department of
JOURNAL Genetics, Washington University School of Medicine, 4444 Forest
Park Boulevard, St. Louis, MO 63108, USA
COMMENT Supported by NIH grant 50 01 A143283

Coding sequences below are predicted from manually evaluated
computer analysis, using similarity information and the programs:
GLIMMER: <http://www.tigr.org/softlab/glimmer/glimmer.html> and
GeneMark: <http://opal.biology.gatech.edu/GeneMark/>

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>,
and Pedro Romero and Peter Karp at EcoCyc;
<http://ecocyc.org/ecocyc/PangeaSystems.com/ecocyc/>

The analyses of ribosome binding sites and promoter binding sites
were kindly provided by Heladia Saigado, Julio Collado-Vides and
Regunondi;
http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate
chemistries or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one m13 subclone.
Location/Qualifiers
1. 26591
/organism="Salmonella typhimurium LT2"
/strain="LT2; SGSC 1412; ATCC 700720"
/db_xref="ATCC:700720"
/db_xref="taxon:99287"
/complement(68. .2238)
/gene="wzc"
/note="STM2116"
complement(68. .2227)
/gene="wzc"
/note="Ortholog of E. coli orf, hypothetical protein
(AAC75121.1); Blastp hit to AAC75121.1 (732 aa), 87%
identity in aa 13 - 732"
/codon_start=1
/transl_table=11
/product="putative tyrosine-protein kinase in colanic acid
export"
protein_id="AAL21020.1"
/db_xref="GI:16420650"
/translation="MTEKAKSAVATGSDIDIGRIVCTEARWVIGTAIPALAA
VIVFAPRISADALVOIKONAGNSIVDTNSALAKPPASDAEIQIISRVLGKT
VDDLDIAVRKNTFPLFGAGWERLMGHNHNVKVFPTPTMSGQIFLTKVIGDKR
YQLVSDGGSAGVVGDPPLNDGVTRVEALDARPDTEFTVSKFSTGLMNNLNLT

gene
/complement(2224. .2685)
/gene="wzb"
/note="STM2117"
complement(2224. .2673)
/gene="wzb"
/EC_number="3.1.3.48"
/note="Ortholog of E. coli probable
protein-tyrosine-phosphatase (AAC75122.1); Blastp hit to
AAC75122.1 (147 aa), 83% identity in aa 1 - 146"
/codon_start=1
/transl_table=11
/product="putative protein-tyrosine-phosphatase in colanic
acid export"
protein_id="AAL21021.1"
/db_xref="GI:16420651"
/translation="MFNKLIVCVGNCVCRSTARLLKRFPSLTVAAGLAVGK
ADPAASVSAHDLSENHGCAROIARLCRYDILMERHITAACDIPBRGKVM
LFGHDSREIHPDPRKSRDPAFEAVYLLERSAQMNALNAEGKP"
complement(2233. .2238)
/gene="wzc"
/note="Putative RBS for wzc; RegulonDB: STMS1H002262"
complement(2679. .3945)
/gene="wza"
/note="STM2118"
complement(2679. .3818)
/gene="wza"
/note="Ortholog of E. coli putative polysaccharide export
protein (AAC75123.1); Blastp hit to AAC75123.1 (379 aa),
93% identity in aa 1 - 379"
/codon_start=1
/transl_table=11
/product="putative polysaccharide export protein, outer
membrane"
protein_id="AAL21022.1"
/db_xref="GI:16420652"
/translation="MMSKKMLPILASLISGCVLPSSNMSTGKXIVIKQADP
DLDWVAVPLTPELVOLRPRNVAPNMSIDOLASVYRGPRDVLVYWDPE
LTPAGVYRSSSGNMVOPDGMFPTVYIGKVSYGKTSIERSDITGRVLTADPQ
VDVNIAPRSQKATISGVNKSQOATITVPLVDLINAAGLITMAMRVNLTN
GKEORISLQALMONGDLSQNRLLYPGDIIVPNNDDIKFVMEVKKOSTLKQDFSGM
LTREALGNAEGIDLTTSNAGSIFVIRPLKGGGGRGKIANIYOLDMSDATSLVATEF
RLQPYDVVYVYTPAVARWRLINOLLPTISGVRYMTDITASDIHSW"
complement(2680. .2685)
/gene="wzb"
/note="Putative RBS for wzb; RegulonDB: STMS1H002263"
complement(3918. .3926)
/gene="wza"
/note="Putative -10_signal for wza;
RegulonDB: STMT1H004524"
complement(3937. .3945)
/gene="wza"
/note="Putative -35_signal for wza;
RegulonDB: STMT1H004524"
4476. .6073
/gene="yegH"
/note="yegH"
/gene="STM2119"
4476. .4481
/gene="yegH"
/note="Putative RBS for yegH; RegulonDB: STMS1H002264"
4493. .6073
/gene="yegH"
/note="Ortholog of E. coli putative transport protein
(AAC75124.1); Blastp hit to AAC75124.1 (549 aa), 90%
identity in aa 23 - 549"

```
/codon_start=1
/transl_table=11
/product="putative inner membrane protein"
/protein_id="AA121023.1"
/db_xref="GI:16420653"
/translation="MEWIDPSIWAGLVTLVIEVLGIDNLVFITALLAEKLPQGRD
RARITGLITAMIMRLILASISWLVTLKPLFVSQALSFSADLIMLGFFLLFKAT
MELNERLEGKDSANPTORAKAFMAVVAQIVYDAIFSLDSVITAVGVDHIAVMAA
VITISIMILASKSLTRFVNNPTIIVICLSFLMIGFSLVAGFGPHIPKGYLYAAI
GSVMIEALNOLATFNRRRLSANHTLRORTTEAVMRILSGKKEDELDAETAMLAD
HDSQITFNPQERRMTERVNLNQTSSIMTSRHDIEHIDANPAEIRALAEKQHT
RLVTVGENEQEDLGVVHVIDLQSLGSEPLNLALIRQPLVPEPILPLPALEOR
NARTFAFVDEFGSVEGIVTLDVMEIAGNLPNEVEIDARDIQKNPDGSWTANG
HMPLEDVQVYPLPDEKREYHTIAGLMEYLQRIPTGEEVQVDYLLKTYLOVESHR
VOKYOLIPHNGEPPQEW"
complement(6159..8024)
/gene="asma"
/nc="STM2120"
complement(6159..8015)
/gene="asma"
/nc="Ortholog of E. coli suppressor of ompF assembly
mutants (AAC75125.1); Blastp hit to AAC75125.1 (617 aa),
76% identity in aa 1 - 617"
/codon_start=1
/transl_table=11
/product="suppressor of ompF assembly mutants"
/protein_id="AA121024.1"
/db_xref="GI:16420654"
/translation="MKRFLITLMLLVIVAGFSALVLLVNPDPFAYMVOQVARSSE
YOLLDGRLRMVHPQLSTLSGRMTLTAGASEPLVRADNMRLDVALMPLSHQLVK
OVALKGVVQLTPTQEAIVSDDAVPAKDNLTLDVAEDGMSFDISLRVADSVLVFQ
HENDDQVTRDITRLNMEQDAHEGTFDESGRVNRDQDLALSGSTVADSDYPNLTIA
GTEQLRWOLGADLPAQGLEGOGLOAOQWQEAOKRLSEFNHNLTFANDSSLTGOVOTL
ARPEWQIDLOSSKLNLDNLPHHSAVAQOTGAVGOGONTLPLTRPVIASRYGAPPYO
GIQSTFAETALADSVIRKGMPTFOYSTMSQAGLIDTIDLOGLADGEMSLPSTLD
ARTASPRIEHPRLNHEVETGTLKAFNPISLTGKSLVGFDSGADIDAEARHSMKG
KAHVDMSTNRLBGMFQOLVQOAVERSGDAQOQSENMNATRLDRFTTDLTKNGTL
TLDMDVGOSSMLALTGSTLDLVKOSCTDFNLRYLGGWNGDSNLITFLKETPVPLRV
YGRWOLNTLYOVDLLRKHLDEAKRRLNDWADRNRKDTNRNGDKVKKLNKL"
complement(8019..8024)
/gene="asma"
/nc="Putative RBS for asma; RegulonDB:STMS1H002265"
complement(8054..8649)
/gene="dcd"
/nc="STM2121"
complement(8054..8635)
/gene="dcd"
/nc="number=3.5.4.13"
/nc="Ortholog of E. coli 2'-deoxycytidine
5'-triphosphate deaminase (AAC75126.1); Blastp hit to
AAC75126.1 (193 aa), 95% identity in aa 1 - 193"
/codon_start=1
/transl_table=11
/product="dtrpase"
```

```
Query Match 67.6%; Score 19.6; DB 1; Length 26591;
Best Local Similarity 84.6%; Pred. No. 9e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 cacgcaagcgcgctcgccagccac 26
||||| ||||| ||||| |||||
DB 8980 CACGCAAGCGCGCTCGGTACGTAC 9005
```

Search completed: July 31, 2002, 14:00:49
Job time: 17491 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 09:06:17 ; Search time 165.21 Seconds
(without alignments)
5627.521 Million cell updates/sec

Title: US-09-899-718a-1

Perfect score: 3785
Sequence: 1 gtttggttcgctgttttc.....tggaccgctgtcgcac 3785

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 38353 segs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	185	4.9	4800	3 US-08-941-445A-4	Sequence 4, Appl1
2	137.8	3.6	2542	3 US-08-941-445A-6	Sequence 6, Appl1
3	118.8	3.1	2267	4 US-08-679-645-25	Sequence 25, Appl1
4	83	2.2	2549	1 US-08-470-720-2	Sequence 2, Appl1
5	83	2.2	4964	1 US-08-470-720-5	Sequence 5, Appl1
6	71.2	1.9	19124	2 US-08-487-826B-13	Sequence 13, Appl1
7	64.8	1.7	7218	1 US-08-232-463-14	Sequence 14, Appl1
8	62.2	1.6	7218	1 US-08-232-463-14	Sequence 14, Appl1
9	58.8	1.6	5852	1 US-07-867-106-2	Sequence 2, Appl1
10	56.6	1.5	658	4 US-08-998-416-595	Sequence 555, App
11	56.4	1.5	5852	1 US-07-867-106-2	Sequence 2, Appl1
12	55.6	1.5	51952	3 US-08-947-823-1	Sequence 186, App
13	54.2	1.4	615	4 US-08-998-416-186	Sequence 1, Appl1
14	53.8	1.4	636	4 US-08-998-416-1137	Sequence 137, App
15	52.4	1.4	837	4 US-08-998-416-288	Sequence 288, App
16	52.2	1.4	19124	2 US-08-487-826B-13	Sequence 13, Appl1
17	51.2	1.4	240	1 US-08-628-417-6	Sequence 6, Appl1
18	51	1.3	1511	1 US-07-991-867B-8	Sequence 8, Appl1
19	51	1.3	1511	1 US-08-107-755A-8	Sequence 8, Appl1
20	51	1.3	1511	2 US-08-544-332-8	Sequence 8, Appl1
21	50.4	1.3	4673	1 US-07-638-431-1	Sequence 1, Appl1
22	50.4	1.3	4673	5 PCT-US93-00018-1	Sequence 10, Appl1
23	50.4	1.3	4766	5 PCT-US93-07261-10	Sequence 10, Appl1
24	50.2	1.3	5181	1 US-08-257-073-10	Sequence 8, Appl1
25	49.6	1.3	3211	2 US-08-574-959A-8	Sequence 8, Appl1
26	49.6	1.3	3211	2 US-09-357-014-8	Sequence 6, Appl1
27	49.6	1.3	3901	2 US-08-574-959A-6	Sequence 6, Appl1

28	49.6	1.3	3901	4 US-09-357-014-6	Sequence 6, Appl1
29	49.6	1.3	71989	4 US-09-443-501A-2	Sequence 2, Appl1
30	49	1.3	6243	2 US-09-056-075-1	Sequence 1, Appl1
31	48.6	1.3	2230	4 US-08-378-313-24	Sequence 14, Appl1
32	48.6	1.3	2447	2 US-09-014-969-14	Sequence 101, App
33	48.6	1.3	43795	3 US-08-742-185-101	Sequence 27, Appl1
34	48.4	1.3	4818	3 US-08-817-926-27	Sequence 3, Appl1
35	47.8	1.3	2997	3 US-08-486-343A-3	Sequence 1, Appl1
36	47.8	1.3	2997	5 PCT-US95-07349-3	Sequence 8, Appl1
37	47.4	1.3	5394	3 US-08-688-376-1	Sequence 1, Appl1
38	47.2	1.2	602	1 US-08-764-100-8	Sequence 13, Appl1
39	47.2	1.2	642	1 US-08-764-100-13	Sequence 7, Appl1
40	47.2	1.2	643	1 US-08-764-100-7	Sequence 2, Appl1
41	47.2	1.2	2993	1 US-08-764-100-2	Sequence 10, Appl1
42	47.2	1.2	2993	1 US-08-764-100-10	Sequence 9, Appl1
43	47.2	1.2	3000	1 US-08-764-100-9	Sequence 1, Appl1
44	47.2	1.2	3001	1 US-08-764-100-1	Sequence 1, Appl1
45	46.6	1.2	51952	3 US-08-947-823-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-08-941-445A-4
; Sequence 4, Application US/08941445A
; Patent No. 6107060
; GENERAL INFORMATION:
; APPLICANT: Keeling, Peter
; TITLE OF INVENTION: Starch Encapsulation
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle
; CITY: Boulder
; STATE: CO
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/941.445A
; FILING DATE: 30-SEP-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,855
; FILING DATE: 30-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Winner, Ellen P
; REGISTRATION NUMBER: 28,547
; REFERENCE/DOCKET NUMBER: 89-97
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4800 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: not relevant
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ORIGINAL SOURCE:
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(1449..1553, 1685..1765, 1860..1958, 2055
; ..2144, 2226..2289, 2413..2513, 2651..2760, 2858
; ..3101, 3212..3394, 3490..3681, 3793..3879, 3977

FILING DATE: 19920625
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PJ 7187
APPLICATION NUMBER: PCT/AU90/00530
FILING DATE: 02-NOV-1989
ATTORNEY/AGENT INFORMATION:
NAME: Feeney, Joanne Longo
REGISTRATION NUMBER: 35,134
REFERENCE/DOCKET NUMBER: RICE-0002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-568-3100
TELEFAX: 215-568-3439
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5852 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 2378..5038
NAME/KEY: CDS
LOCATION: 2378..5038
US-07-867-106-2

Query Match 1.5%; Score 56.4; DB 1; Length 5852;
Best Local Similarity 48.1%; Pred. No. 0.0011;
Matches 228; Conservative 0; Mismatches 236; Indels 10; Gaps 2;

OY 314 taagtattatcacccattttgcagtcataaattatgcatttcgtacaaattgctg 373
DB 5549 TTAATAATTATGATCGATTTTAAATCAATGACGATTAATAAATCTTAAAC 5490
OY 374 gcaactctcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 433
DB 5489 AATAAACATATTTGATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 5430
OY 434 gatactaattatgcctcatttggaatttcgttttggaattatgctag------ta 485
DB 5429 AATAAT 5370
OY 486 cacacttctctgtatattatggaagcgcaatttcctgtatagtttcttctgt 545
DB 5369 AAAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5310
OY 546 atttttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctg 605
DB 5309 ATTATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5250
OY 606 cttagaataacttagtatttgatttgatttgatttgatttgatttgatttgattt 665
DB 5249 CCAAGTTTTCCTTTTATTTATTTATTTCTCTCATGTTTCTTATGAGCTTTTCT 5192
OY 666 aagggaataacacatgcacactaatccattcattcattcattcattcattcattc 725
DB 5191 CCAGGCAAAATTAACAGATCAAGAAATCTCATTAATTTTGGTCTCAAAAATTTTAA 5132
OY 726 actcaacttattgcttattgcttattgcttattgcttattgcttattgcttattg 779
DB 5131 TCTCATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5078

RESULT 12
US-08-947-823-1
Sequence 1, Application US/08947823
Patent No. 6114605
GENERAL INFORMATION:
APPLICANT: Williamson, Valerie M.
APPLICANT: Kaloshian, Isouhi

APPLICANT: Yaghoobi, Jafar
APPLICANT: Bodeau, John
APPLICANT: Milligan, Stephen
TITLE OF INVENTION: Procedures and Materials for Confering
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESS: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/947,823
FILING DATE: 09-OCT-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US97/18802
FILING DATE: 09-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/028,191
FILING DATE: 10-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Bastian, Kevin L.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 023070-07021005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 51952 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-947-823-1

Query Match 1.5%; Score 55.6; DB 3; Length 51952;
Best Local Similarity 43.9%; Pred. No. 0.005;
Matches 238; Conservative 0; Mismatches 304; Indels 0; Gaps 0;

OY 125 tcatgtttatttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 184
DB 12960 TAATGTCCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 13019
OY 185 acttaagggaatttcagtttatttactagtttcagtttatttactagtttatttact 244
DB 13020 TGGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 13079
OY 245 tttagttggttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 304
DB 13080 TTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 13139
OY 305 atatacaataaglatatatacacccatttttgcagtcataaaattatgcaatttcag 364
DB 13140 TACTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 13199
OY 365 aaatttggtgcaactcttcttcttcttcttcttcttcttcttcttcttcttcttctt 424
DB 13200 AAACCTTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 13259
OY 425 ataacaatgataactaattatgctcatttgggaatttcgttttggaattatgctag 484
DB 13260 AACCCCTACCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 13319

TELEPHONE: 919-541-8587
 TELEFAX: 919-541-8689
 INFORMATION FOR SEQ ID NO: 1137:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 636 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ORIGINAL SOURCE:
 ORGANISM: PAGI692RP
 US-08-998-416-1137

Query Match 1.48; Score 53.8; DB 4; Length 636;
 Best Local Similarity 45.98; Pred. No. 0.0016;
 Matches 266; Conservative 0; Mismatches 302; Indels 11; Gaps 2;

```

QY 196 aattcagtttatactagttcagtttattatgatttattagtttattagttt 255
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 58 ATTAGTATTAATAATTATTAATACTATTAATCATATTATTAATAATTATTTGAT 117

QY 256 tctcattatgtagcaagaatataagggtgtgtgtgtgtgtgtgtgtatatacaca 315
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 118 TATTAACTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 177

QY 316 agtatatacccaattttgcagtcataaataagcaatttcagtcagtcagtcagtc 375
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 178 TATTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 237

QY 376 aaactctctcatttatttatttatttatttatttatttatttatttatttatttatt 435
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 238 TAGTATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 297

QY 436 tactaattatgcctcatttgcgaattcgttttgaaaa-----ttatgctagtaga 487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 298 CAATTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 357

QY 488 caattatctgtatataattgaaaaagcgaatttctgtgtaagt---ttgtcattctg 544
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 358 GATTTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 417

QY 545 tatttttttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 604
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 418 TATTAGTGAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 477

QY 605 gcttagaagaactttagtattttagtattttagtattttagtattttagtattttagt 664
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 478 ACTTAACTTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 537

QY 665 taagggaatacacaatgcacaaatccatccatccatccatccatccatccatccatcc 724
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 538 TTATTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 597

QY 725 aaactcaactttatagcttattcgtgcataattataaa 763
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 598 AAAATATTAATAAGATGAGTAAAAAATCTTATATAAA 636
  
```

RESULT 15
 US-08-998-416-288
 Sequence 288, Application US/08998416
 Patent No. 6239264

GENERAL INFORMATION:
 APPLICANT: Philippsen, Peter
 APPLICANT: Pohlmann, Rainer
 APPLICANT: Steiner, Sabine
 APPLICANT: Mohr, Christine
 APPLICANT: Wendland, Jurgen
 APPLICANT: Knechtel, Philipp
 APPLICANT: Reibschung, Corinne
 TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPYII
 TITLE OF INVENTION: AND USES THEREOF

NUMBER OF SEQUENCES: 1152
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: No. 6239264artis Corporation
 STREET: 3054 Cornwallis Road
 CITY: Research Triangle Park
 STATE: No. 6239264th Carolina
 COUNTRY: USA
 ZIP: 27709
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/998,416
 FILING DATE: 24-DEC-1997
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: CH 0016/97
 FILING DATE: 31-DEC-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Meigs, J. Timothy
 REGISTRATION NUMBER: 38,241
 REFERENCE/DOCKET NUMBER: PF/5-30306/A/GC1976
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8587
 TELEFAX: 919-541-8689
 INFORMATION FOR SEQ ID NO: 288:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 837 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ORIGINAL SOURCE:
 ORGANISM: PAGI241RP
 US-08-998-416-288

Query Match 1.48; Score 52.4; DB 4; Length 837;
 Best Local Similarity 44.68; Pred. No. 0.0039;
 Matches 258; Conservative 0; Mismatches 311; Indels 9; Gaps 1;

```

QY 196 aattcagtttatactagttcagtttattatgatttattagtttattagttt 255
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 58 ATTAGTATTAATAATAATAATACTATTAATCATATTATTAATAATTATTTGAT 117

QY 256 tctcattatgtagcaagaatataagggtgtgtgtgtgtgtgtgtgtatatacaca 315
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 118 TATTAACTTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 177

QY 316 agtatatacccaattttgcagtcataaataatgcgaatttcagtcagtcagtcagtc 375
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 178 TATTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 237

QY 376 aaactctctcatttatttatttatttatttatttatttatttatttatttatttatt 435
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 238 TAGTATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 297

QY 436 tactaattatgcctcatttgcgaattcgttttgaaaa-----ttatgctagtaga 486
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 298 CAATTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 357

QY 487 aaactctctgtatataattgaaaaagcgaatttctgtgtaagtattgtaattctgta 546
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 358 GATTTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 417

QY 547 ttatttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 606
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 418 TATTAGTGAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 477

QY 607 ttagaagaactttagtattttagtattttagtattttagtattttagtattttagtatt 666
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  
```

D_B 478 ACTTTAAATTCCTTATTATTAATTTTTATAATTAATTANAATTAAATTAATTGCATTTTATTTA 537
QY 667 agggaaatacaaacagccacatcatcattccaactgagaanaatcttlaaccttaaccaaaaa 726
D_B 538 TTTATTATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 597
QY 727 ctcaacttttatatcgctatccgcgatcataataaaa 764
D_B 598 TAAAATTATTATTAAGAAATGTGATGAATAAATACCTATTA 635

Search completed: July 31, 2002, 12:17:49
Job time: 11492 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 08:27:27 ; Search time 5855.71 Seconds

(without alignments)
8724.130 Million cell updates/sec

Title: US-09-899-718a-1

Perfect score: 1 gtttggttcgcgtgttttc.....tgacgcggtgttcgcgcac 3785

Sequence: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Scoring table: 13736207 segs, 6748477542 residues

Searched: Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST :
1: em_estha :
2: em_esthum :
3: em_estin :
4: em_estnu :
5: em_estov :
6: em_estpl :
7: em_estro :
8: em_estc :
9: gb_estl :
10: gb_estl2 :
11: gb_hic :
12: gb_gss :
13: em_gss_hum :
14: em_gss_inv :
15: em_gss_pln :
16: em_gss_vrt :

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	325.8	8.6	527	9	AM448831 BRY_1580
2	293.4	7.8	579	10	BE402419 CSB007G05
3	293.2	7.7	558	10	BE602243 HVSMEH09
4	291.6	7.7	638	9	AL508959 AL508959
5	283.8	7.5	639	10	BE414303 SCU008.CO
6	278.8	7.4	700	9	AL508163
7	258.8	6.8	513	9	AL508933
8	248	6.6	677	10	BE414500
9	247	6.5	700	9	AL506567
10	231.4	6.1	444	10	BM368238
11	220.6	5.8	494	9	AM448811
12	210.2	5.6	462	9	AL506396
13	178.4	4.7	375	9	AM448845
14	118.6	3.1	1885	10	BE420745
15	114	3.0	1135	12	CNS0336Q
16	112	3.0	879	12	CNS01JRG
17	111.8	3.0	470	10	C73133

C	18	111.4	2.9	1036	12	CNS03LWJ	AL250012 Tetraodon
	19	110.4	2.9	1059	12	CNS0022B	AL097133 Drosophila
	20	110.2	2.9	907	12	CNS021J4	AL176953 Tetraodon
	21	109.4	2.9	855	3	BI645015	BI645015 OP2815 M1
	22	109.4	2.9	896	10	BM415641	BM415641 OP20719 M
	23	106.4	2.8	1198	12	B08337	B08337 T19P9-SP6.1
C	24	106.2	2.8	1101	12	CNS012JN	AL101645 Drosophila
	25	105.8	2.8	1317	10	BE420688	BE420688 HMM001.F0
	26	105.4	2.8	964	10	BM416130	BM416130 OP21217 M
C	27	104.6	2.8	1101	12	CNS0153V	AL104965 Drosophila
	28	104.4	2.8	1101	12	CNS00K85	AL077453 Drosophila
	29	104.2	2.8	982	12	AQ325799	AQ325799 nxb0021B
	30	104	2.7	834	12	BI2387	BI2387 F21E20-SP6.
	31	104	2.7	1184	12	BI3117	BI3117 T9K3-SP6.2
C	32	103.6	2.7	614	12	CNS0152H	AL104915 Drosophila
	33	103.2	2.7	960	12	AC031602	AC031602 Pan trogl
	34	102.8	2.7	848	3	BI644518	BI644518 OP2257 M1
	35	102.8	2.7	870	12	AQ330286	AQ330286 nxb0046J
	36	102.6	2.7	1139	12	AQ897537	AQ897537 HS.3153.A
	37	102.6	2.7	1223	12	BI2981	BI2981 T24D11-SP6
	38	102.4	2.7	924	10	BM415947	BM415947 OP21029 M
	39	102.2	2.7	1185	10	BE273407	BE273407 CA_EB001
	40	102	2.7	656	12	CNS01ZDA	AL174367 Tetraodon
	41	101.8	2.7	1309	10	BE420736	BE420736 HMM002.A0
	42	101.6	2.7	622	12	CNS04ROH	AL304226 Tetraodon
	43	101.6	2.7	966	10	BM415686	BM415686 OP20766 M
	44	101.2	2.7	865	12	AQ324474	AQ324474 mgb0018B
C	45	101	2.7	855	12	AZ183849	AZ183849 SE_1002.A

ALIGNMENTS

RESULT 1
LOCUS AM448831 527 bp mRNA linear EST 03-JAN-2001
DEFINITION BRY_1580 BRY Trilicium aestivum CDNA clone p56-1P, mRNA sequence.
ACCESSION AM448831
VERSION AM448831.1 GI:12019366
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Trilicium aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Trilicaceae; Trilicium.
1 (bases 1 to 527)
REFERENCE Clarke,B.C., Hobbs,M. and Appels,R.
Genes active in developing wheat endosperm
Unpublished (2000)
JOURNAL Contact: Bryan Clarke
COMMENT Division of Plant Industry
C.S.I.R.O.
GPO Box 1600, Canberra, ACT, Australia
Tel: 61 2 6246 5054
Fax: 61 2 6246 5000
Email: bryan@pi.csiro.au.

FEATURES
source location/Qualifiers
1..527
/organism="Trilicium aestivum"
/cultur="Wyna"
/db_xref="taxon:4565"
/clone="P56-1P"
/clone_id="BRY"
/cell_type="endosperm"
BASE COUNT 115 a 176 c 163 g 73 t
ORIGIN

Query Match 8.6%; Score 325.8; DB 9; Length 527;
Best Local Similarity 99.4%; Pred. No. 5.5e-38;
Matches 327; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

07 3092 atcagtcagtcgtctctgtgtcaggtacacacccctgcgcgcgcgtgcgtctg 3151

QY 3472 cc 3473
 Db 530 CC 531

RESULT 7
 AL508933 513 bp mRNA linear EST 04-JAN-2001
 LOCUS AL508933 Hordeum vulgare Barke developing caryopsis (3.-15.DAP)
 DEFINITION Hordeum vulgare cDNA clone HY10D16V 5', mRNA sequence.
 ACCESSION AL508933
 VERSION AL508933.1 GI:12035436
 KEYWORDS EST.
 SOURCE
 ORGANISM Hordeum vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 Triticeae; Hordeum.
 1 (bases 1 to 513)
 ; Triliceae; Hordeum.

REFERENCE
 AUTHORS Michael, W., Meschke, W., Pleissner, K.-P. and Graner, A.
 TITLE EST sequencing and analysis in barley
 JOURNAL Unpublished (2000)
 COMMENT Contact: Michael W
 Institute for Plant Genetics and Crop Plant Research
 Corrensstr.3, D-06466 Gatersleben, Germany
 Email: michael@ipk-gatersleben.de, http://pgrc.ipk-gatersleben.de
 Seq primer: T3 primer for 5' end.
 Location/Qualifiers
 1. 513
 /organism="Hordeum vulgare"
 /cultivar="Barke"
 /db_xref="taxon:4513"
 /clone_1id="HY10D16V"
 /clone_1lb="Hordeum vulgare Barke developing caryopsis
 (3.-15.DAP)"
 /tissue_type="developing caryopsis (3.-15.DAP)"
 /lab_host="X10LR"
 /note="Vector: Plasmid pBK-CMV; Site_1: EcoRI; Site_2:
 XhoI; mRNA was made from developing caryopsis (3.-15.DAP)
 of spring barley variety 'Barke', a high quality milling
 variety. Cloning sites: EcoRI (5'-end of cDNA) and XhoI
 (3'-end of cDNA). NOTE: Due to a cloning artefact caused
 by the kit, in most cases the EcoRI site is NOT present,
 as well as the EcoRI adapter. Average insert size is 1 kb
 trimmed from the 5'-end until a 50 bp window
 contains less than two ambiguities. The maximum length was
 set to 700 bp"

BASE COUNT 102 a 184 c 151 g 73 t 3 others
 ORIGIN

Query Match 6.8%; Score 258.8; DB 9; Length 513;
 Best Local Similarity 90.9%; Pred. No. 3.1e-28;
 Matches 298; Conservative 0; Mismatches 24; Indels 6; Gaps 2;

QY 3099 cagtcgtctcttctcgtcagtagcacacccctgcgc---gcgcacatgcygcctctgtgca 3155
 Db 188 CAGTCGCTCTCTACGTGACGTACACACCTGTGCGGCGCCGACATGGCGGCTCTGCGCA 247

QY 3156 cctccacagctcgcacactcgcgcacgtctcagtgtaaccgcgaattccggcgtccag 3215
 Db 248 CGTCCAGCTCTGCGCAGCTCTGCGCAGCTCTGCGCAGCTCTGCGCAGCTCTGCGCAG 307

QY 3216 gtttcaaggcctcagagcccggaacccgcgagatcgcgctcgcgcatgaagacgtctg 3275
 Db 308 GTTTTCAAGGCGCTCAGAGCCCGGACCAACCGGATGCGGCGCTGTGATGAGAGACTATCG 367

QY 3276 gaacgagcgccgcccacaaagcaagcaaacgcacgcatcgacgcgagtgctctct 3335
 Db 368 GACCAAGCGCCGCGCCGGAAGCAAGCCGACCGCGGAGCGCGGAGCGGCTGCTCT 427

QY 3336 ccattggtggtggtgcgcgcacagcgcgagcgagtagtaactgtgtctcgtcgcgcgaga 3395
 Db 428 CCGTGTGTGTGCGCGCCACGCGCA---GCGCATGAACCTGTGTGTGCGCGCGGAGA 484

QY 3396 tggcgcccttgagcaagacatgctgcgcct 3423
 Db 485 TGGCGCCCTTGAGCAAGACCGCGCGCT 512

RESULT 8
 BE414500 677 bp mRNA linear EST 24-JUL-2000
 LOCUS BE414500 SCU010.D11.R990714 TREC SCU Wheat Endosperm Library
 DEFINITION aestivum cDNA clone SCU010.D11, mRNA sequence.
 ACCESSION BE414500
 VERSION BE414500.1 GI:9412346
 KEYWORDS EST.
 SOURCE
 ORGANISM Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 Triticeae; Triticum.
 1 (bases 1 to 677)
 ; Triliceae; Triticum.

REFERENCE
 AUTHORS Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T., Cloutier
 S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A., Gustafson, P.,
 Herrmann, R.G., Holton, T., Jacquemijn, J.M., Jia, J., Joudrier, P.,
 Langridge, P., Lazo, G.R., Lin, J.J., McGuire, P., Ogihara, T.,
 Pecchioni, N., Qaiset, C., Schuch, W., Selvaraj, G., Shariflou, M.,
 Sorrells, M., Warburton, M. and Wenzel, G.
 International Triticaceae EST Cooperative (TREC): Production of
 Expressed Sequence Tags for Species of the Triticeae
 Unpublished (2000)
 COMMENT Contact: Holton T
 Centre for Plant Conservation Genetics, Southern Cross University
 PO Box 157, Lismore NSW 2480 AUSTRALIA
 Tel: 61 2 6620 3409
 Fax: 61 2 6622 2080
 Email: tholton@scu.edu.au
 International Triticaceae EST Cooperative (TREC)
 http://wheat.pw.usda.gov/genome.
 Location/Qualifiers
 1. 677
 /organism="Triticum aestivum"
 /cultivar="Wynna"
 /db_xref="taxon:4565"
 /clone_1id="SCU010.D11"
 /clone_1lb="TREC SCU Wheat Endosperm Library"
 /tissue_type="endosperm"
 /note="Vector: Bluescript II SK(-)"

BASE COUNT 138 a 207 c 212 g 97 t 23 others
 ORIGIN

Query Match 6.6%; Score 248; DB 10; Length 677;
 Best Local Similarity 83.5%; Pred. No. 1.1e-26;
 Matches 313; Conservative 0; Mismatches 57; Indels 5; Gaps 3;

QY 3092 atcagtgagtgctctctgtcgtcagtagcacacccctgcgcgacatgagcgctctg 3151
 Db 254 ATCAACCAAGTGTCTCTGCTGCGCAGTAGCACACCTGTGCGGCGCGCATGCGGCTCTG 313

QY 3152 gtacagctccagctcgcacactcgcgcacgtctcagtcagctcagcagacagattccgct 3211
 Db 314 GTACAGTGTGCGCAGCTCTGCGCAGCTCTGCGCAGCTCTGCGCAGCTCTGCGGCT 373

QY 3212 ccaggtttcagagcctcagagcccggaacccgcgagatcgcgctcgcgcatgaagact 3271
 Db 374 GCAAGTTTTCAGGCGTGTAGGCGCCCGGAGCCCGGACATGCGCGCTGTGAGGACT 433

QY 3272 gtccgagagagccgcgcacca---aagcaagcaggaagaaaccgacgcatccgagcg 3328
 Db 434 ACCGAGAGGAGCGCGCCCGCGAAGCAACAAAGCCCGAAGGCGACCCGCGGAGCGCGG 493

Thu Aug 1 08:29:57 2002

us-09-899-718a-1.1st

01 Zea mays.
02
03 PN WO9844780-A1.
04 XX
05 PD 15-OCT-1998.
06 XX
07 PF 03-APR-1998; 98WO-US06660.
08 XX
09 PR 04-APR-1997; 97US-0042939.
10 XX
11 PA (EXSE-) EXSEED GENETICS LLC.
12 XX
13 PI Guan H, Keeling PL;
14 XX
15 DR WPI: 1998-568285/48.
16 XX
17 DR P-PSDB; AAY16604.
18 XX
19 PT Producing non-glycogen-like polysaccharides in bacteria, fungi or
20 PT plants - transformed with genes for enzymes involved in starch or
21 PT glycogen synthesis allows fermentative production of starches with
22 PT engineered properties
23 XX
24 PS Disclosure; Fig 49; 150pp; English.
25 XX
26 CC The specification describes a method for the production of
27 CC non-glycogen-like polysaccharides in a host. The method comprises
28 CC transforming a host, suitable for fermentation, with genes encoding
29 CC starch- or glycogen-synthesis enzymes, and fermenting the transformants.
30 CC The specification also describes hosts transformed with a gene active
31 CC in glycogen synthesis and at least one non-starch branching gene,
32 CC involved in production of amylopectin or amylose in its original host.
33 CC The method is used to produce plant-like starches by fermentation and
34 CC new starches in plants. These starches are useful for all food and
35 CC non-food applications of starch. The present sequence is used in
36 CC the course of the invention.
37 XX
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
76

QY	464	cgtttgaaaatagctagtcacactatcttctgtacataatgaagaagcgcaatttc	523
Db	3107	tt	3166
QY	524	tgtgtgaagtcttgcatcctgtaattttttccattctctcctcgtgaagggtaacac	583
Db	3167	cgctttg	3226
QY	584	taatgcacaactaattcattctgttagaaaaacttagtatctgttgtttagtttt	643
Db	3227	tt	3286
OY	644	tattcatcttgcttctcttt	665
Db	3287	tttgttttgttttttttttttt	3308
RESULT 14			
ID	AAS46686	standard; DNA: 7442 BP.	
XX	AC	AAS46686;	
XX	DT	18-DEC-2001 (first entry)	
XX		Tumour suppressor gene derived chemically modified sequence #409.	
DE		Human; tumour suppressor gene; oncogene; antitumour; cytostatic;	
KW		cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;	
KW		cytosine methylation; ds.	
XX			
OS		Homo sapiens.	
XX			
PN		WO200168912-A2.	
XX			
PD		20-SEP-2001.	
XX			
PF		15-MAR-2001; 2001WO-EP02955.	
PR			
XX		15-MAR-2000; 2000DE-1013847.	
PR		06-APR-2000; 2000DE-1019058.	
PR		07-APR-2000; 2000DE-1019173.	
PR		30-JUN-2000; 2000DE-1032529.	
PR		01-SEP-2000; 2000DE-1043826.	
XX			
PA		(EPTG-) EPIGENOMICS AG.	
PI		Olek A., Piepenbrock C, Berlin K;	
PI			
DR		WPI; 2001-602752/58.	
XX			
PT		Fragments of chemically modified genes associated with tumour suppressor	
PT		genes and oncogenes, useful in designing primers and probes for	
PT		analysing diseases associated with cytosine methylation state e.g.	
PT		cancer	
XX			
PS		Claim 1; SEQ ID NO 409; 27pp; English.	
XX			
CC		The invention relates to a nucleic acid comprising a sequence of 18	
CC		bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with	
CC		bisulphite, of genes associated with tumour suppression and	
CC		oncogenes having a sequence taken from 536 (actually 533 since	
CC		numbers 408, 458 and 500 are missing from the sequence listing) sequences	
CC		(SS) and sequences complementary to (SS). The nucleic acid may be a	
CC		peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may	
CC		form part of a set of probes for detecting the cytosine methylation state	
CC		and/or single nucleotide polymorphisms and also to be used in an	
CC		array for analysing diseases associated with CpG dinucleotides e.g.	
CC		cancers and tumours. The probes can also be used in a method for	
CC		ascertaining genetic and/or epigenetic parameters for the diagnosis	
CC		and/or therapy of existing diseases or the predisposition to specific	
CC		diseases, by analysing cytosine methylations. The parameters may be	

CC compared to another set of genetic and/or epigenetic parameters, the
CC differences serving as basis for diagnosis and/or prognosis events which
CC are disadvantageous to patients. The present sequence is one of the
CC 533 genomic sequences derived from tumour suppressor genes and
CC oncogenes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).
XX
SQ Sequence 7442 BP; 1655 A; 214 C; 1934 G; 3636 T; 3 other;

Query Match	2.6%	Score 98.6;	DB 22;	Length 7442;
Best Local Similarity	46.8%	Pred. No. 4.2e-12;		
Matches 31;	Conservative	0;	Mismatches 354;	Indels 0;
				Gaps 0;

Qy	1	gtttggttcgcgtgttttcaattccctctctcctaagggaataccaaatgcagtaat	60
Db	1909	gttttcaatttttagattttttttttaaatttttttttttagattttttttttttta	1968
Qy	61	tcaatattgtaacagcgcgcattctcgtgcacaataatgtaacattctcttgtaatg	120
Db	1969	tttttttttttttttagagtcgcgttttttttttttttttttttttttttttttt	2028
Qy	121	tggttaagtgtttattcaattctcttcaactttttagggtaaaacaabccccaatca	180
Db	2029	tgtgttttttttttttagaatttttttttttttttttttttttttttttttttt	2088
Qy	181	ttcacccaagagaagaatcagtttatatactagtttcagtttatattgtttataatg	240
Db	2089	tttaatttttttttttagatttttttttttttttttttttttttttttttttt	2148
Qy	241	gtttttagttggtttcccatltagtaatgtaacgaatacaggggtgtgtgcggt	300
Db	2149	tttttcaattttttaaatttttttttttttttttttttttttttttttttttt	2208
Qy	301	gttaataacacataagatataacaccatttttgacgcaataaaatcgaattca	360
Db	2209	tttatcttttttttttagatttttttttttttttttttttttttttttttttt	2268
Qy	361	gtacaattgtgcgcacaactctcttaattttttttttttttttttttttttctttaa	420
Db	2269	tattttttttttttaatttttaatttttttttttttttttttttttttttttttt	2328
Qy	421	gggttaacaacaatgatactaaattttgcacatttggaaattcgtttggaaaaatgac	480
Db	2329	tattttttttttttaatttttttttttttttttttttttttttttttttttt	2388
Qy	481	tagtaacaactattctgttaattatgaaggaagcgcaatttcgtgtgaagtttgcatt	540
Db	2389	tatttttttttttttagaatttttaatttttttttttttttttttttttttttttt	2448
Qy	541	tctgatttttttttcaatttctctctctgtggaaggttaacaactaaatgcactaatcat	600
Db	2449	tt	2508
Qy	601	tctgtctagaagaacttgatattttgattgtgtgttttaatttttttttttttttttt	660
Db	2509	tttttttttttttagattttatttttttttttttttttttttttttttttttttt	2568
Qy	661	tttt	720
Db	2569	ttttt	2573
RESULT 15			
ABL32527			
ID ABL32527 standard; DNA: 7571 BP.			
XX	ABL32527;		
AC			
XX			
26-MAR-2002 (first entry)			

Thu Aug 1 08:29:56 2002

us-09-899-718a-1.rng

Page 14

Query Match 100.0%; Score 3785; DB 6; Length 3785;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3785; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtttggttcgtgttttcaattccttcttcttaagggttaatacaatgaaat 60
DB 1 GTTGGTTGCGTGTTCATTTCCTTCTTAAAGGGGTAAATACCAATGACAGTAAT 60
QY 61 tcaatttgtaacagtgagatctcttgccaattatgtaacattctcttgtaatt 120
DB 61 TCATATTGTGTAACTGCGATCTTGAGCAATTATGACATTTCTTTGTATTTGT 120
QY 121 ttttcaattcattcattcttcttactctttaggttaaaacaatgcccacaat 180
DB 121 TGTTCATGTTTATTTATTTATTTCTTTAGGTAACCAATGCCCAATTTCA 180
QY 181 tttaactaagaagaatcagatttataactagttcaagttattattattgtaagt 240
DB 181 TTTTACCTAAGAGAAATTCAGTTTATTAACAGTTTATTAATTTATTTATTAAGT 240
QY 241 gtttttagtggtttctcattatgtaatgtaataataggggtgtgtgtgtgt 300
DB 241 GTTTTATGTTGTTTCTCATTTATGTAATGATGAATATGAGGGGTGTGTGCTGT 300
QY 301 gtaataacaacataagatataacacacatttttcagtcataaataatgcaattca 360
DB 301 GTTATATATACACATATGATATATACACCAATTTTGGACATCAATTAATTCATCA 360
QY 361 gtaacaatttgccgaacactctctcattcttcttcttcttcttcttcttctaa 420
DB 361 GTACAAATTTGGCCAAACCTCTTCTTCATTTTATTTTATTTATTTCTTCTTAA 420
QY 421 gggtaataccaatgataactaattatgcccatttggaatttcgttttgaaatagc 480
DB 421 GGGTAATACCAATGATACATTTATATGCTCATTTGGAAATTCGTTTGAATAATTAAGC 480
QY 481 tagtaacaactatcttgaatataatgaaagcgaattctcgtgaattgtgtat 540
DB 481 TAGTACACACTTATCTTGATATATGAAAGGCAATTTCTGTATAGTTTGTAT 540
QY 541 tctgtatttttttcaatttcttctctctctggaagggttaacaataatgccaattatca 600
DB 541 TCTGATATTTTTCATTTTCTTCTCTGGAAGGTAACACTTAATGCAATATGAT 600
QY 601 tctgtctagaagaacttagatattgtatgtgttcttatttatttcttcttct 660
DB 601 TCTGCTTAGAAGAACTTATGATTTTGTATTTGTTTATTTATTTTCTTCTTCT 660
QY 661 tctttaagggaataacaaatgccaactcaatcattcattagaaatctcttattctta 720
DB 661 TCTTTAAGGAAATACCAATGACCAATTCATTCATCTTAAGAAATCTCTTAATCTTA 720
QY 721 caaaaactcaacttttatalgcttalcgtgataatataaaagcagttcttactcta 780
DB 721 CAAAAACCTCACTTTTATGCTTATTCGTGATATTAATAAAGGCAAGTCTTCTATCTTA 780
QY 781 aattgctgcaaaacttatacatattgtcttaataatttttctcagaatgtaattacc 840
DB 781 AATTGGCTGCAAACTTTATATCAATTTTGTCTAAATTAATTTTCTGAAATGATGATAC 840
QY 841 aatgccaactaattcattccgtgagcagcaatatacggaatgccaatgataatagt 900
DB 841 AATGCCACTAATTCATTCCTGAGCAAGCAATATGAGGATGCGTATTAATTAAGT 900
QY 901 gtgcatttttcaatctcaacgcatgggcatgataccctcaacatgcaacacagcat 960
DB 901 GTGCATTTTTCATCTCTACGCAATGGGCAATGCAATACCATGACACACAGCAAT 960
QY 961 aacaaacaatagacactcaacgagagcaatgataacactggttgcgaacaaagacag 1020
DB 961 ACACAACATGAGCACTCACGCGACATGCAATACCTGTGTGGACACACACACAC 1020

QY 1021 cgacacacagcagcagccacatgctgtgcaacttagaagaataaataagacgtatacaatt 1080
DB 1021 CGACACACAGCCACAGCCACATGCGTGTGCACTTTAAGAAAAAATATGACACTATACATT 1080
QY 1081 ggaactgtaactatactaccgtgtaaacactagtaactgtgtgtgtgtgtgtgtgtgt 1140
DB 1081 GGACTGGTACTACTACTACTACCGGTACATAGTACGTAGTGTGTGTGTGTGTGTGTGT 1140
QY 1141 aagtgccaagaactagatatttcaaggcgaactgggataagcagcgcattgtgtgtgt 1200
DB 1141 AGGTGCCACACATGATATTTTTCAGCGCATGCGGATATAGCCACAGGCTATTTGTTCTGT 1200
QY 1201 tctgtagaacgaagaacggtatataatgtgtaactgtgcttcttaagaactcacaaggct 1260
DB 1201 TCTGTAGACGAAACGGTCTATATGTGTGCACTGTGCTCTTATAGAGACTCTCCAAAGGCT 1260
QY 1261 caacacactacacgtgagtgacagcccaacgctgcgtgtaaaaccaacgcaatttaagttccc 1320
DB 1261 CACACACTCACCGTGTAGTGTACAGCCCGCTGCTTAACACCGCATTTTATGCTTTTCC 1320
QY 1321 gatcgaacaaagcgaaggaacagcagtaactgttcaatgttgacagctgtgtgtgtgt 1380
DB 1321 GATCGACAAAGCCAGGAGCAAGCACTGCTGCTCATGTTGTGTGTGTGTGTGTGTGTGT 1380
QY 1381 acgagcaggttggcagcagcagtaactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1440
DB 1381 ACGCGCGGTTTGGAGACAGTACGTGCTGCTTATACCAAGCCGTATGCTCATATCA 1440
QY 1441 agcaaaagaagaagaagaaggcgaagaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1500
DB 1441 AGCAAAAGAGAAAAAGAGGGGGGAAAGGTGATAGCCCGGCTGTGTGTGTGTGTGTGTGT 1500
QY 1501 aggaagaacatcccggtgcatgacgcatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1560
DB 1501 AGGAACAATTCGGGGCCATGCAAGCGCATGTGCAAGCGCCCAAGCAAAAGCAAGCGAG 1560
QY 1561 agcgaagacacacatggt 1620
DB 1561 AGCGAGAGCAGCATAGGCGCCCACTGAAACCGAGGAGACACAGAAAGCGCGGTGCT 1620
QY 1621 gctgtgaactgaacaggaagaac 1680
DB 1621 GCGTGAACATCACACAGAAACACCCACCGGACCGCCAGCGCGCGCGCGCGCGAGAC 1680
QY 1681 agaagatgctgtcagcagcgt 1740
DB 1681 AGAAGATGGGTGACAGCGCGCGCGCGCAAGGAGGCGCGCGCGCGCGCGCGCGCGCGCG 1740
QY 1741 aaagcggt 1800
DB 1741 AAAGCGCGGT 1800
QY 1801 caagctgtgcaagcgt 1860
DB 1801 CAGGCTGTGCAAGCGGT 1860
QY 1861 cgtctggtgcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1920
DB 1861 CGCTGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1920
QY 1921 ctcacactcaacgt 1980
DB 1921 CTCACACTCAACGT 1980
QY 1981 tctatgacatgtgagcccaacagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2040
DB 1981 TCTATGACATGTAGCCCAACAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2040
QY 2041 cctttaagacacgaagcgt 2100
DB 2041 CCTTTAAGACCGAAGT 2100

[illegible]

Db	3181	CCTGTCAGCGTACCGACAGATTCCGGCGCTCCAGGTTTCAGGGCTTGAGGCCCCGGAA	3240
QY	3241	ccggcgagatgcggtcgtctcggaattgaggactgttcgagcgcgcgccccaagaacaa9	3300
Db	3241	CCGGCGGATGC GGCGGTCCGGCATGAGAGACTGTTCGAGAGCAGCGCGCCCCAAGCAAG	3300
QY	3301	caggaaaccgcaccgatcttgaccggcggtgcctctcaatggtgtgtgcgcgcgcggcg	3360
Db	3301	CAGGAAMCCGCACCAGATTCCACCGCGGgtccttccatgtgtgtgcgcgcgcgcgcg	3360
QY	3361	cggcgagatgaacctcgctggttcgttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	3420
Db	3361	CGCGCGCATGAACCTGTGTTCGTTCGCGCGCGCGGATGCGCCCTGGAGCAAGACTGGCGG	3420
QY	3421	cctcggcgagttcctctgggggtcctccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	3480
Db	3421	CCTCGCGAGGTCCTCGGGGGGCTCCCGCGGCATGCGCGTAAGCTTCGCCACATCGCTT	3480
QY	3481	tcttaataatgttcttcttcgc	3540
Db	3481	TCTTATAATAGTTTCTTCCTCGACGCCATGCTTCCTTCACACGGGTGCCGTTCCTGTC	3540
QY	3541	aggcacaagtcacgcgggtcatgtcatctcccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	3600
Db	3541	AGGCCAACGGTACCGGGTCATGTCATTCCTCCCGCGCTACGCACAGTACAAGSAGCGCT	3600
QY	3601	gggacacacgc	3660
Db	3601	GGGACACACGCGTCANCTCCGAGGTATATATCCGCCACATGAAATATACAAATTCACATG	3660
QY	3661	ctctcgacattctcgcgaacttiactgactggtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	3720
Db	3661	CTCTCTCACATTTCTGCAAGACTTTACTGATGGGTGATCTCGAGATCAAGATGCTGG	3720
QY	3721	aagggtacagagggtgagtactcttcaactgtctaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	3780
Db	3721	ACGGGTACAGAGGGTAGTCTTCACATGCTTACAAAGCGCGGGTGAGCCGCTGTTCG	3780
QY	3781	tcgac 3785 	
Db	3781	TCGAC 3785	
<hr/>			
RESULT	2		
AB019622			
LOCUS	AB019622	2805 bp	DNA linear
DEFINITION	Triticum aestivum gene for starch synthase (GBSSI), complete cds.		PLN 05-AUG-1999
ACCESSION	AB019622		
VERSION	AB019622.1 GI:4760579		
KEYWORDS	starch synthase (GBSSI).		
SOURCE	Triticum aestivum DNA.		
ORGANISM	Triticum aestivum		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;		
AUTHORS	Poideae; Triticeae; Triticum. 1 (sites)		
TITLE	Murai,J., Taira,T. and Ohta,D. Isolation and characterization of the three Waxy genes encoding the granule-bound starch synthase in hexaploid wheat		
JOURNAL	Gene 234 (1), 71-79 (1999)		
MEDLINE	99321800		
REFERENCE	2 (bases 1 to 2805)		
AUTHORS	Murai,J., Taira,T. and Ohta,D.		
TITLE	Direct Submission		
JOURNAL	Submitted (10-NOV-1998) Jun Murai, Osaka Prefecture University, College of Agriculture; 1-1 Gakuen-cho, Sakai, Osaka 599-8531, Japan (E-mail:junki@edemeter.plant.osaka-fu-u.ac.jp, Tel.:81-722-54-9409, Fax:81-722-54-9409)		
FEATURES	Location/Qualifiers		
SOURCE	1..2805		
	/organism="Triticum aestivum"		
	/db_xref="taxon:.4565"		

gene

join(13..333,416..496,581..679,789..942,1068..1168,
1268..1621,1713..1892,1988..2179,2270..2356,2455..2583,
2677..2793)
/gene="waxy"

CDS

join(13..333,416..496,581..679,789..942,1068..1168,
1268..1621,1713..1892,1988..2179,2270..2356,2455..2583,
2677..2793)
/gene="waxy"

/codon_start=1

/product="starch synthase (GBSSI)"

/protein_id="BA077350.1"

/db_xref="GI:4760580"

/translation="MAALVTSQLATSGTVLSTDRFRFRPGGLRPNPADAALGMR
/VGAAPKOSRKHPRFRDRCLSMVVRATGSGGMNVPGAMAPMSKGTGLDYLGL
PAAAMANGHVRVISPREDQYKADMDTVSIEIKVADYERFRFHCYKRGVDFVD
HPCLEKVRGKTKERIKYDAGTDEYEDNOQRFSLCOALEPRILDLNNPHSGPY
GSDVYFVCDMDHTGLACYSKNSQNSGIIQADKVLVSPYAEELISGARGLD
PDRKSSDFIDGDKPYEGRKINMKKAGIIOADKVLVSPYAEELISGARGLD
NIMRLTGTIVNGMDVSEMDPIKDKPLTVNDVTALGKALKEALQAEVGLPDR
KTVLAFRLEBOKGPDYMTAIPETVEEDVOILVLTGKPKFERLKSVEKEFT
KVAVAFRNPALHQMAGADVLATVSPREPGLIOLQGMRYGTCACASGLVDTI
VSKTGFHNGRLSDCNVNEPADVKRVVITLKRKVVGTIPATYHEMVKNCMIQDLSMK
GPAKMEVDLLELGVESGEPGIVGEIAPLALENVAAP"

BASE COUNT

577 a 817 c 826 g 585 t

ORIGIN

Query Match

17.4%; Score 658; DB 8; Length 2805;

Best Local Similarity 100.0%; Pred. No. 2.5e-115; Mismatches 0; Gaps 0;

Matches 658; Conservative 0; Indels 0;

QY 3128 cctcgccgagccatcgagcgtctgctacgtccacagctccacacccgacgctctc 3187

Db 1 CCTCGCGCGCCAGAGGGGCGCTGTGACGTCACGTCGCCACCTCCGGCACCCTCTC 60

QY 3188 agcgtcacccagacgattccgagcgtcttcaagggccttgagggcccggaacccggcg 3247

Db 61 AGCGTACCGACAGATTCCGGCGTTCAGGGCTGAGGCCCGCGAACCCCGCGC 120

QY 3248 gatcgagcgtcgagcgtgagcgtgagcggcgcccccgaagcaagcaggaaa 3307

Db 121 GATGGGGGCTCGGCATCAGAGCTGTGGAGCGAGCGCCGCCCAAGCAAGAGGAAA 180

QY 3308 ccgacacgattccgacccgagcgtgctctcctcctcctcctcctcctcctcctc 3367

Db 181 CCGCACCCATTCCGACCGCGGCTGCTCTCCATGCTGTGGCGCGCCAGCGGAGCGCGC 240

QY 3368 atgaacctcgtctcgtcgtcgagcggcgcccttgaggaagcgtgagcgtcgcc 3427

Db 241 ATGAACCTCGTGTCTGTCGGCGCGAGATGGCGCCCTGAGACCAAGACTGGGGGCTCGGC 300

QY 3428 gacgctcccgagggggtcccgccgacatggcctgaaagcttgagcgtcctctata 3487

Db 301 GACGCTCCGGGGGCTCCCGCGCGCATGGCCCTGAAGCTTGGCCGACCTGCTTATA 360

QY 3488 aatgcttctcctgacgacatgctcgtttacaaggggtgctgctgctgagcgcaa 3547

Db 361 AATGTTTCTCTCTGACGACATGCGCTGCGGTTACAAAGGGGTCCGTCGTCAGGCCAA 420

QY 3548 cgggtacccgggtatgctatcctcccgctcgaagcgtacaaagcgtcgtggagac 3607

Db 421 CCGTACACGGGTCATGTCATCTCCCGCTACACACAGTAAAGAGCGCTCGGACAC 480

QY 3608 cagcgctatctccgaggtatataccgcaatgaattacaattacaattcgtcctcgc 3667

Db 481 CAGCGCTATCTCCGAGGTATATATCGCCACATGATTAATATCAATTAATCAATGCTCGC 540

QY 3668 acatttctgcaagacttactgactgagctcgcagatcaagagtcgttgacagta 3727

Db 541 ACATTTCGCAAGACTTACTGACTGGCTGATCTCGCAGATCAAGGTGCTGTGACAGGTA 600

QY 3728 cgagaggtgaggtactcactgctacaagcggtgagcggtgagcggtgctgctgac 3785

|||||

Db 601 CGAGAGGTGAGTACTTCCACTGCTACAAAGCGCGGGGTGACCGCGTGTCTGCAC 658

RESULT 3

AB029061 2781 bp DNA linear PIN 27-JUN-2001
LOCUS titlicum subsp. diococcoides waxy gene for starch synthase
DEFINITION (GBSSI), complete cds.

ACCESSION AB029061

VERSION 1

KEYWORDS starch synthase (GBSSI).

SOURCE titlicum subsp. diococcoides

ORGANISM titlicum subsp. diococcoides

REFERENCE 1 (sites)

AUTHORS Murali, J., and Ohta, D.

TITLE Isolation and characterization of the four waxy genes encoding the

JOURNAL granule-bound starch synthase in tetraploid wheats

REFERENCE Appl. Biol. Sci. (1999) in press

AUTHORS Murali, J., and Ohta, D.

TITLE Direct Submission

JOURNAL Submitted (16-JUN-1999) Jun Murali, Osaka Prefecture University,

College of Agriculture, 1-1 Gakuen-cho, Sakai, Osaka 599-8531,

Japan (E-mail: junkiedemeter.plant.osakaifu-u.ac.jp,

Tel: 81-722-54-9409, Fax: 81-722-54-9409)

FEATURES Location/Qualifiers

1..2781

/organism="Triticum turgidum subsp. diococcoides"

/sub_species="diococcoides"

/db_xref="taxon:85692"

join(1..321,404..484,569..667,777..930,1056..1156,

1256..1609,1701..1880,1976..2167,2258..2344,2443..2571,

2655..2781)

/gene="waxy"

join(1..321,404..484,569..667,777..930,1056..1156,

1256..1609,1701..1880,1976..2167,2258..2344,2443..2571,

2655..2781)

/gene="waxy"

/codon_start=1

/product="starch synthase (GBSSI)"

/protein_id="BA08509.1"

/db_xref="GI:6624281"

/translation="MAALVTSQLATSGTVLSTDRFRFRPGGLRPNPADAALGMR

/VGAAPKOSRKHPRFRDRCLSMVVRATGSGGMNVPGAMAPMSKGTGLDYLGL

PAAAMANGHVRVISPREDQYKADMDTVSIEIKVADYERFRFHCYKRGVDFVD

HPCLEKVRGKTKERIKYDAGTDEYEDNOQRFSLCOALEPRILDLNNPHSGPY

GSDVYFVCDMDHTGLACYSKNSQNSGIIQADKVLVSPYAEELISGARGLD

PDRKSSDFIDGDKPYEGRKINMKKAGIIOADKVLVSPYAEELISGARGLD

NIMRLTGTIVNGMDVSEMDPIKDKPLTVNDVTALGKALKEALQAEVGLPDR

KTVLAFRLEBOKGPDYMTAIPETVEEDVOILVLTGKPKFERLKSVEKEFT

KVAVAFRNPALHQMAGADVLATVSPREPGLIOLQGMRYGTCACASGLVDTI

VSKTGFHNGRLSDCNVNEPADVKRVVITLKRKVVGTIPATYHEMVKNCMIQDLSMK

GPAKMEVDLLELGVESGEPGIVGEIAPLALENVAAP"

BASE COUNT

569 a 811 c 818 g 583 t

ORIGIN

17.0%; Score 642.8; DB 8; Length 2781;

Best Local Similarity 99.7%; Pred. No. 2e-112;

Matches 644; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3140 atggcggctctgltacagtcctcagcagcctccgacacgctcaggtacacgac 3199

Db 1 ATGGCGGCTCTGTGTCAGCTCCGACCTCCGCGCACCGCTCTCAGCTCACCGAC 60

QY 3200 agatctcggcggtcaggtttcagggcctgagggcccggaacccggcgatcgcgctc 3259

Db 61 AGATTCCGGGCTCCAGGTTTTCAGGGCTGAGGCCCGCGAACCCGCGGATCGGCGCTC 120

QY 3260 ggcattgagactctcgagcgagcgcccccgaagcaagcaggaacccgacacgac 3319

|||||

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
SOURCE

barley.
Hordeum vulgare
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooidae; Triticeae; Hordeum.
1 (bases 1 to 124050)
Ma,J., Samikunel,P.J., Dubcovsky,J., Shiloff,B.A., Rostoks,N.,
Jiang,Z., Busso,C.S., Kleinholts,A., Devos,K.M., Ramakrishna,W. and
Benneken,J.L.
Comparative sequence analysis of wx1 homologous regions in barley,
maize, pearl millet, rice, sorghum and diploid wheat
Unpublished
2 (bases 1 to 124050)
Ma,J., Samikunel,P.J., Dubcovsky,J., Shiloff,B.A., Rostoks,N.,
Jiang,Z., Busso,C.S., Kleinholts,A., Devos,K.M., Ramakrishna,W. and
Benneken,J.L.
Direct Submission
Submitted (25-JUN-2002) Department of Biological Sciences, Purdue
University, West Lafayette, IN 47907, USA
Location/Qualifiers
1..124050

misc_feature
156..3499
/note="similar to putative copII-like retroelement pol
polyprotein".
join(<6912..6697,7144..7232,7460..7562,7713..7773,
8287..8367,8772..8832,8944..>8975)
/gene="259116.1"
/product="putative leucine aminopeptidase"
<6912..>8975
/gene="259116.1"
join(<6912..6697,7144..7232,7460..7562,7713..7773,
8287..8367,8772..8832,8944..8975)
/gene="259116.1"
/codon_start=1
/product="putative leucine aminopeptidase"
/protein_id="FAL77104.1"
/db_xref="GI:18652402"
/translation="MGSGTMRKAGALNDSTFVGLTNTLSTPNNFSGLDDIIATIVA
TNHVECPKKRHRRTSVNPPHILCYKPGVGNVKKRLATVIGLTFPDSGGYNIKTGP
GSGICTMRARRRDNINMGRGLTADALVYACNGCVYDKIIDLTLGACVVALPSPJA
GLTFPSDELA"
join(<10203..10391,11938..>12408)
/gene="259116.2a"
/product="unknown"
<10203..>12408
/gene="259116.2a"
join(<10203..10391,11938..12408)
/gene="259116.2a"
/codon_start=1
/product="unknown"
/protein_id="FAL77105.1"
/db_xref="GI:18652403"
/translation="NNHAEKRYKDAASNAKAKITQAKVAEKTEATARSHDERELA
HRRGAKVAALAEHLQAKITFARVAKSATSNEIDAVDEAVGKLAVIDTSGEBGAA
TIKSEALANQAGVAGVGRGLVLLAASHDOEATLDEVVGHGRRRRPMWGTAAL
TSGRSATVPLDRLPLCFSSLLTAELCMAPGGLGRVHVYQGLIQSHMGKNQKRRKKNR
K"
join(<15600..15852,16734..>16939)
/gene="259116.2b"
/product="unknown"
/protein_id="FAL77106.1"
/db_xref="GI:18652404"
/translation="NNHAEKRYKDAASNAKAKITQAKVAEKTEATARSHDERELA
HRRGAKVAALAEHLQAKITFARVAKSATSNEIDAVDEAVGKLAVIDTSGEBGAA
TIKSEALANQAGVAGVGRGLVLLAASHDOEATLDEVVGHGRRRRPMWGTAAL
TSGRSATVPLDRLPLCFSSLLTAELCMAPGGLGRVHVYQGLIQSHMGKNQKRRKKNR
K"
join(<15600..15852,16734..16939)
/gene="259116.2b"
/codon_start=1
/product="unknown"
/protein_id="FAL77106.1"

repeat_region
/db_xref="GI:18652404"
/translation="MNIAKKEVKDAASAkakATITAAKVAEKTEATARSHDEELA
HERGAKAKAAAEALYQAKVTHEERAMEHRLHKRHSHSGVDILINALLFPVNLCCYON
VETGVDEHETIQLAVSLIGVADLTIVHKGISTADDCSGGIEKSTGG"
16128..16358
/transposon="MITE"
join(<18744..19118,21680..>22075)
/gene="259116.3"
/product="hypothetical protein"
<18744..>22075
/gene="259116.3"
join(18744..19118,21680..22075)
/gene="259116.3"
/codon_start=1
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="AI177108.1"
/db_xref="GI:18652405"
/translation="MDASLLPEVTLTEVAHASPRLDVIAGVCAVQAQASAVPARKQ
GNVYVGGHPAPAKARAPGAVVVSRRPTAEKYTKAGVKGFWLHPLSPSKRE
MPKRVSTSTLEPRKRYGRNKKALCGAGDDEELRLCDICDGRFTFLRGS
APMPPEPSMSPATOSRGLGEDELHGARVRLMLRSPPLSLVDYSTGLEGSGEAR
NSWKGDESGRGGCGRDGVVQVTEGACARNKRAMP"
20238..20342
/transposon="MITE"
/complement(23173..26525)
/note="similar to putative non-LTR retrotransposon"
28545..37482
/transposon="LTR-retrotransposon BARE-1"
complement(37482..46140)
/transposon="LTR-retrotransposon BARE-1"
46161..47870
/note="similar to mudra-like transposon"
complement(57099..64304)
/note="repetitive sequence Rp-1"
/evidence=experimental
58591
/rpt_unit=eg
58592..58663
/rpt_unit=at
64444..64565
/transposon="MITE"
complement(64715..69626)
/transposon="LTR-retrotransposon Sukkula_solo"
65014..65035
/rpt_unit=gaa
<70375..>70971
/gene="259116.4"
/product="hypothetical protein"
<70375..>70971
/gene="259116.4"
70375..70971
/gene="259116.4"
/codon_start=1
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="AA177108.1"
/db_xref="GI:18652406"
/translation="MPKRRSASGIRGVRRPMPGGYAIRSGDRLSLGTYDPAHEA
ARADVAAMWLGRLRQINFDQVHTLQADLVAIPRVASADRDHTRARORLLVA
EDEYVAAMKRRHEDVAVEQGYWARRREEDTRRREELDRRRKRALASAHADIVA
GGRSEFTFEEDERFWFIDWISDTDDDDGVDWSDMD"
72677..72709
/rpt_unit=gaa
join(<72864..73181,73270..73350,73435..73533,73660..73813
73950..74050,74165..74518,74611..74790,74885..75076,
75159..75245,75343..75471,75548..>75664)
/gene="259116.5"
/product="granule-bound starch synthase"
/evidence=not_experimental
72864..75664
/gene="259116.5"
join(72864..73181,73270..73350,73435..73533,73660..73813,
73950..74050,74165..74518,74611..74790,74885..75076,
73950..74050,74165..74518,74611..74790,74885..75076,

Db	Accession	Length	Score	DB	Length
Db	481 TG6GACACCA6CGTCGTCAGACTTGAACCCCTAACCTTAAACATCAAA	540	13.1%	Score 495.2	DB 8
Qy	3653 ttgcacatgcctcctgcaacattctgcaagacttaactgactg----gctgatatcgcaga	3708	89.4%	Pred. No. 3e-84	
Db	541 TTCGCAATGCTCTGCAATATTTCTGCAGAAATCTACTGACTGACTTAATGATCTCCGAGA	600			
Qy	3709 tcaagatcgttgacaaggtaagagaggttgaggtacttcacactgctctcaaacgcgggggtg	3768			
Db	601 TCAAGTCGTTGACAAATACAGAGAGGSGTAGGACTTCTCACTGCTACAGACGCGGGGtgG	660			
Qy	3769 acgcgctgctcgtcgtcgcac 3785				
Db	661 ACCGCGTGTCTGTCGAC 677				
RESULT	10				
LOCUS	AB029064	2804 bp	DNA	linear	PLN 21-DEC-1999
DEFINITION	Triticum durum waxy gene for starch synthase (GBSSI), complete cds.				
VERSION	AB029064.1	GI:6624286			
KEYWORDS	Starch synthase (GBSSI).				
SOURCE	Triticum durum DNA.				
ORGANISM	Triticum durum subsp. durum				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Triticum.				
AUTHORS	1 (sites)				
TITLE	Murai,J., Taira,T. and Ohta,D.				
JOURNAL	Isolation and characterization of the four Waxy genes encoding the granule-bound starch synthase in tetraploid wheats				
REFERENCE	Appl. Biol. Sci. (1999) In press				
AUTHORS	2 (bases 1 to 2804)				
TITLE	Murai,J., Taira,T. and Ohta,D.				
JOURNAL	Submitted (16-JUN-1999) Jun Murai, Osaka Prefecture University, College of Agriculture; I-1 Kakuen-cho, Sakai, Osaka 599-8531, Japan (E-mail: junkiedometer.plant.osakafu-u.ac.jp, Tel:81-722-54-9409, Fax:81-722-54-9409)				
FEATURES	Location/Qualifiers				
source	1..2804				
gene	/organism="Triticum durum subsp. durum"				
	/db_xref="taxon:4567"				
	join(1..321,418..498,590..688,799..952,1093..1193,1293..1646,1731..1910,1993..2184,2265..2351,2439..2567,2688..2804)				
CDS	/gene="waxy"				
	join(1..321,418..498,590..688,799..952,1093..1193,1293..1646,1731..1910,1993..2184,2265..2351,2439..2567,2688..2804)				
	/gene="waxy"				
	/codon_start=1				
	/product="starch synthase (GBSSI)"				
	/protein_id="BAA8512.1"				
	/db_xref="GI:6624287"				
	/translation="MAALVTSQLATSGTILGTTDFRRAGFGVGRPNPADALVMTIGASAAQKSRKARRGSRCLSMVVRATGSGNMLVVGAEAMPSKGTGLGVLGELPAAANQRVAVISPRDOYKQDAMDITSSVEIRVADYEERVFHCYKRGVDVFPDHPCLLEKVRGKTEKIKIYCPDAGTGTEEDNDQLEFSLICQALAEPIILDLNNPVSFGYGEDVAVFCNDKHTGILACYSKTSVNSQNGITWTAVCINHSIQGRFSPDFRQMLNPMRLSGTIVNGMDVSEMDPTKDFLAVNYDVTTALDEGALKKELQLQAEVGLPDKVPLVAFILGRLEEGKGPDMVTALPELIKEDVOYLILGKKKFERLLSVSEKFPNKKRVAVRRNAPLAAHQMAADVLAVTFRPSCGCIQLQGRAYGPCACASTGGLVDTIVEGTGTGMRGLSYDCNVVEPADAVKRVFTTLTKRAVKVGPAYHGVKNCMIDLSMKGPANMEVLLLELGEVSESPVIEELTAPLAMEVNAAP"				
BASE COUNT	577 a 823 c 824 g 580 t				
ORIGIN					

College of Agriculture, 1-1 Gakuen-cho, Sakai, Osaka 599-8533,
Japan (E-mail: junkie@nemer.plant.osaka-fu-u.ac.jp,
Tel: 81-722-54-9409, Fax: 81-722-54-9409)

QY 3641 gaattacacattcacaatcgtctcgtcacattctcgaagacttactgactggtgat 3700
 Db 538 GAATATACAGAGTTCACATGCTCCTCCACATTTCTGCAGAGTTCACACTACCGAGTGGAT 597
 QY 3701 ctgcagatcaaggtgctgtgacaggtgacagaggtgaggtactctcactgctacaagcg 3760
 Db 598 TTCACAGATCAAGTTCGCGGACGAGTACGAGGAGTGTACTTCCACTCTACAGAGG 657
 QY 3761 cggagtgacacgcgtgtctgacac 3785
 Db 658 CGGGGTGACCGCGTCTTCTGCTGAC 682
 RESULT 12
 AF110374 2826 bp DNA linear PLN 12-NOV-1999
 LOCUS Aegilops speltoides granule-bound starch synthase wx-Tsb protein
 DEFINITION (wx-Tsb) gene, complete cds.
 ACCESSION AF110374
 VERSION AF110374.1 GI:6318539
 KEYWORDS Aegilops speltoides.
 SOURCE Aegilops speltoides.
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Poideae; Triticeae; Aegilops.
 REFERENCE 1 (bases 1 to 2826)
 AUTHORS Yan, L., Bhawe, M., Fairclough, R., Konic, C., Rahman, S. and Appels, R.
 TITLE The genes encoding granule-bound starch synthases at the waxy loci
 of the A, B and D progenitors of common wheat
 JOURNAL Genome (1999) In press
 REFERENCE 2 (bases 1 to 2826)
 AUTHORS Yan, L., Bhawe, M., Fairclough, R., Konic, C., Rahman, S. and Appels, R.
 TITLE Direct Submission
 JOURNAL Submitted (02-DEC-1998) School of Life Sciences and Technology,
 Victoria University of Technology, Werribee Campus, PO Box 14428
 MCMC, Melbourne Victoria 8001, Australia
 FEATURES
 source
 Location/Qualifiers
 1..2826
 /organism="Aegilops speltoides"
 /cultivar="AUS 21638"
 /db_xref="taxon:4573"
 join(<32..355,455..535,625..723,837..990,1116..1216,
 1316..1669,1734..1933,2016..2207,2291..2377,2466..2594,
 2710..>2826)
 /gene="wx-Tsb"
 <32..>2826
 /note="waxy"
 join(32..355,455..535,625..723,837..990,1116..1216,
 1316..1669,1734..1933,2016..2207,2291..2377,2466..2594,
 2710..2826)
 /gene="wx-Tsb"
 /note="GBSSI"
 /product="granule-bound starch synthase wx-Tsb protein"
 /protein_id="AF06937.1"
 /db_xref="GI:6318540"
 /translation="MAALVTSQATSATVLTGIDRRFHAQFQVRSPADAPLQMR
 VGASAPKQSRRAHGRTRCLSMVVRAGSGMNVFVGAEMAPSKGGIGDFG
 LPPAAMNGHRVVISPRDYDKDAMDTSVSKYADDERRYRHKCRKRDYRVF
 DHPCLFVKYKTKRKETIGDADCTEDNDLRSLCQAALEPRILDLNNPESGP
 YGSDVYFVCDMRTGLLACTKSNVSNQISYTAFCIVSPYAEELISGARGEL
 LPRKSPFEDIDYDKPVEGRKINMKAGILOADVLTSPVYAEELISGARGEL
 DNTMLRGLTIGLVNGADSEMDPTKDNFLAANNITTALEGALEKALQAEVGLPD
 RKVLAFLGRLEQKGPVYMASIPEIKSEEDVQVLLGTGKKRKEALKVEKEFP
 SKYAVVRENAPLAHQMAGADVLAVTSPEEDVQVLLGTGKKRKEALKVEKEFP
 IVEGKTGEHGRSLSDVCNVEPADVKKVYTTLLKRAKVVCTAYHEMVAACMIDLSW
 KGPARKMEVDVLELVEGSEPFYTGSEIPLAMENVAAP
 BASE COUNT 581 a 845 c 815 g 585 t
 ORIGIN

Query Match 12.6%; Score 478.2; DB 8; Length 2826;
 Best Local Similarity 87.2%; Pred. No. 5.3e-81;
 Matches 613; Conservative 0; Mismatches 63; Indels 27; Gaps 7;
 QY 3109 ttgtcgcaggtgagcacaacccctgacgacatgagggcctctgttaacgctccagctcgc 3168
 Db 1 TTGCTGCAGAGTAGGCACACCTTGTGCGGACCATGGCGGCTGTGTACGTCACGATCGC 60
 QY 3169 caactcgcgcacgcgtctcctcagcgtcaccagacaagatctcgcgcgtcgaagcttcaaggcct 3228
 Db 61 CACCTCCGCCACCGCTCTCGGCATCACCAGAGGTTCGGCATGACAGTTTCCAGAGGCGT 120
 QY 3229 gaggcgccggaacccgcgcgcatgctgcgcctcgcgcatagagatgctgcgagagcgccgc 3288
 Db 121 GAGGCCCGGGAGCCCGGACAGATGGCGGCTCGGCATGAGGACGTCTCGACGAGCCCGC 180
 QY 3289 ccccaag---caagcagagaacccgcgcgcatcgcgacccgcgcgttgcctcctcagtggt 3345
 Db 181 CCCGACGACACAAAGCCGGAAGGCGACCCGGGACCCGGCGGTGCTCTGATGGTGT 240
 QY 3346 ggcgcgcagcgagcgagcgagcgtgcatgaaacctgctgttctgcgcgcagagatgagccctg 3405
 Db 241 GCGCGCCACCGGCGAGCGGCGGATGAACTCGTGTGTCGGCGCGAGATGGCGGCTG 300
 QY 3406 gagcagacgtgagcgctcgcgcagctcctcgcgcgcgcctccgcgcagtcgagta-- 3463
 Db 301 GAGCAGACCGGCGGCTCTCGGACGCTCTCGGCGGCTCTCCCGACGATGGCCCTTAAG 360
 QY 3464 -----agcttcgcacactcctctctataa--atgttctcctcgcacatgacgtcgc 3515
 Db 361 CTAGCTAGCTAGACACACCTCTCTCTCATATATTTCTTCTTGCAGGCACCATGCTTGC 420
 QY 3516 -----gtacaacgggtgcgttcctcgtgcagcgaacggtacacggtatccaggt 3568
 Db 421 ATTACAGTAGTACAACTGATGTGTGTGTGTCAGGCAACGTCACCGGATGATGTCAT 480
 QY 3569 ctcccgcgctacgacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3625
 Db 481 CTCGCCCGCTACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 540
 QY 3626 -ataatcgcacacatgaattat--cacaattcacatgctcctcgcacattctgcagaac 3682
 Db 541 CATATATCCCGCCACATGAAATCATCCCAAAATTCATGCTCTCGACATTTCTGCAAGT 600
 QY 3683 ttactgactgctggtatctcgcagatacaagtcgcttgacaggttacgagaggtgagta 3742
 Db 601 TCTACTACCGACTGAT--TTGCAGATCAAGGTCGCCGATGATGACGAGAGGTGAGGTA 659
 QY 3743 ctccactgctacaagcgcgaggtgacgcgctgttcgtcgcag 3785
 Db 660 CTTCAATGTGTACAAAGCGGCGGTGACCGCGGTGTTCTGTCGAC 702
 RESULT 13
 AB029062 2793 bp DNA linear PLN 27-JAN-2001
 LOCUS Triticum turgidum subsp. dicoccoides waxy gene for starch synthase
 DEFINITION (GBSSI) complete cds.
 ACCESSION AB029062
 VERSION AB029062.1 GI:6624282
 KEYWORDS starch synthase (GBSSI).
 SOURCE Triticum turgidum subsp. dicoccoides.
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Poideae; Triticeae; Triticum.
 REFERENCE 1 (sites)
 AUTHORS Mural, J., Talra, T. and Ohta, D.
 TITLE Isolation and characterization of the four Waxy genes encoding the
 granule-bound starch synthase in tetraploid wheats
 JOURNAL Appl. Biol. Sci. (1999) In press
 REFERENCE 2 (bases 1 to 2793)

AUTHORS Murali, J., Taira, T. and Ohta, D.
 TITLE Direct Submission
 JOURNAL Submitted (16-JUN-1999) Jun Murali, Osaka Prefecture University,
 College of Agriculture, 1-1 Gakuen-cho, Sakai, Osaka 599-8531,
 Japan (E-mail: junkiedmeter.plant.osakaifu-u.ac.jp,
 Tel:81-722-54-9409, Fax:81-722-54-9409)

FEATURES

SOURCE

1. 2793
 /organism="Triticum turgidum subsp. dicoccoides"
 /sub_species="dicoccoides"
 /db_xref="taxon:85692"
 join(1..324,424..504,593..691,805..958,1092..1192,
 1262..1615,1708..1887,1974..2165,2249..2335,2433..2561,
 2677..2793)
 /gene="waxy"
 join(1..324,424..504,593..691,805..958,1092..1192,
 1262..1615,1708..1887,1974..2165,2249..2335,2433..2561,
 2677..2793)
 /codon_start=1
 /product="starch synthase (GBSSI)"
 /protein_id="BAA8510.1"
 /db_xref="GI:6624283"

gene

CDS

BASE COUNT 570 a 839 c 820 g 564 t
 ORIGIN

Query Match

Best Local Similarity 87.8%; Score 469.8; DB 8; Length 2793;
 Matches 591; Conservative 0; Mismatches 52; Indels 30; Gaps 6;

QY 3140 atggcgctcgtgtaagtcacagctcgcacactcgcacacgctccacgctcaccgac 3199
 DB 1 ATGGCGCTCTGTGCTACGCTGCACACTCGCACCTCGGACCGTCTCGGCATCACGAC 60
 QY 3200 agattccgagctcagagcttttcaagcgcttgagcccggaacccggcgagatgcgagctc 3259
 DB 61 AGGTTCCGGCGTGCAGGTTTTCAGAGGTGTGAGGCCCGGAGCCGCGCAGATGCCGCTC 120
 QY 3260 ggcataaggaactgtctgagagcgagcgcccccagaag---caagcgaggaacccgagcga 3316
 DB 121 GGCATGAGGACTACCGGAGCGAGCGCCGCCCAACAAAGCCGGAAAGCGCACCG 180
 QY 3317 ttgacagcgagctgctcctcatgctggtgagcgagcgagcgagcgagcatgaactc 3376
 DB 181 GGAACCCGGGCTGCTCTTCATGTTGTCGCGCACGCGAGCGCGGAGTGAACCTC 240
 QY 3377 ggttcgagcgagcagagatggcgccctgagcaagaactgtcgagcctcgagcgagctc 3436
 DB 241 GTGTTGCTGGCGCGGAGATGGCCCTGGAGCAAGACCGGCGCTCGGAGCAGCTCTC 300
 QY 3437 gggagcgctcccgccgagcatgagcgata-----agcttcgagcactgctcttataa 3488
 DB 301 GGGAGCGCTCCCGCCACGATGGCGTGAAGCTAGCTAGCACACCTGTTTGTGAT 359
 QY 3489 atgttcttctcagcagcatgctgcgc-----gttacaagcggtgcgagctc 3536
 DB 360 ATGTTTCTTCTTTCAGACGCAAGCATGCTGCAAGTTTACAAGTGAAGTGAAGTGAAG 418
 QY 3537 gtgcagcgcaagcgagcagcgagctgagctcctcccgagctcagcagcatgaagagc 3596
 DB 419 -TGCAGGCCAAAGGCTACCGGATCATGATCTCCCGCGCTACGACCAAGTGAAGGAC 477

QY 3597 gcttgagacaccagcgatcctccgaggt----atatccgcaatgaattatcaaa 3652
 DB 478 GCTGGAGACACAGCGTCTCCGAGGTACACATATATCCGACATGATATATACAG 537
 QY 3653 ttcaatgctctcagcaattctgcgaagccttactgagctgagctcgagctcgagataa 3712
 DB 538 TTCACTGCTCTCGACATTTCTGCAAGGTTCCACGACCGACTGGATTTCACAGATCA 597
 QY 3713 gtcgttcagagtaagagaggttgagtaactcactgcttacaagcgaggtgagcag 3772
 DB 598 GGTGGGAGCAGAGTACAGAGAGGTGAGTACTTCCACTGCTAACAAGCGCGGTGAGCCG 657
 QY 3773 cgtgttcgagcag 3785
 DB 658 CGTGTCTGTCGAC 670

RESULT 14

TMAXYSS
 LOCUS 2186 bp mRNA linear PLN 02-AUG-1991
 DEFINITION Wheat waxy mRNA for granule-bound starch synthase.
 ACCESSION X57233
 VERSION X57233.1 GI:21901
 KEYWORDS starch synthase; waxy gene.
 SOURCE bread wheat.
 ORGANISM Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Poideae; Triticeae; Triticum.
 1 (bases 1 to 2186)
 Clark, J.R.
 Direct Submission
 Submitted (04-DEC-1990) J.R. Clark, WYE COLLEGE, UNIVERSITY OF
 LONDON, DEPT OF BIOCHEMISTRY & BIOLOGICAL SC., NR. ASHFORD KENT
 TN25 5AH, U K
 2 (bases 1 to 2186)
 Clark, J.R., Robertson, M. and Ainsworth, C.C.
 Nucleotide sequence of a wheat (Triticum aestivum L.) cDNA clone
 encoding the waxy protein
 Plant Mol. Biol. 16 (6), 1099-1101 (1991)
 MEDLINE
 Erratum: [(published erratum appears in Plant Mol Biol 1991
 Oct;17(4):957)]
 Location/Qualifiers
 1. 2186
 /organism="Triticum aestivum"
 /db_xref="taxon:4565"
 /chromosome="7A, 7B, 4B"
 /clone="pCS522"
 /cell_line="Chinese spring"
 /tissue_type="endosperm"
 /clone_id="lambda gt10"
 1..2186
 /gene="waxy"
 /evidence=experimental
 1..2186
 /gene="waxy"
 64..1911
 /gene="waxy"
 /EC_number="2.4.1.11"
 /codon_start=1
 /product="glycogen (starch) synthase"
 /protein_id="CAA40509.1"
 /db_xref="GI:21902"
 /db_xref="SWISS-PROT:P27736"
 /translation="MALVTSQATSTVLTGTRRRRAGQVRRPSPADAPLGMRT
 VGASAPKQSRKARHRTGRLSMVAVATSSAGMNVFVGAEMAPMWSKTGGLGVLDGL
 PAMAMNGHRYVVISPRDQKDMPTSVISEIKYVDYRERYRHYKRGYRVD
 HPCFLKVRGKTEKIKYGPAGDYEDNQRSLCOALEPRILDLNNPFSQY
 AMFLCAVPRAGEYVVCNDMHTGLACYLKNYSQSGIYRTAKVAFCHNTSYGR
 FSHDPAOLNIDRFSSEPFIDGYRPEVGRKIMMKAGIQLADKVLVSPYAAEL
 ISEANGCELDNIMRLGITGIVNMDVSBMPIDKDFLVNVDYVTLALGKALNREA

FEATURES

source

LOAEGVLPVRRKPLVAFIGLEBQKGPDMIAAIPETVKEEDVOIYLLGKKKKFER
 LKSVKEKFTPTVRAVVRNAPLAHOMAGADYLAATVRFEPCCGILIOGMRGTPCA
 CASTGLVDTIVEGTFMGRSLVDNCNVEPADKKVYTLTKRAKRVYGPATHEWY
 KCMIODSLMMKPAKNMEDVLELGVESSEGIYGEETAPLALENVAP"

BASE COUNT 467 a 638 c 694 g 387 t
 ORIGIN

Query Match 9.7%; Score 366.2; DB 8; Length 2186;
 Best Local Similarity 95.4%; Pred. No. 1.2e-59;
 Matches 377; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 3092 atcaatgacgtctgtcttctgtgacgtgacacacccctgcgcgcacatgagcgcctctg 3151
 DB 16 ATCAGACAGTGTCTCTTCTGTCAGGTAGCCACACCCCTGCGCGCCATGCGGCTCTG 75
 QY 3152 gtcaatgccacgtccacacccctgcgcgcacgtctcagcgttaccacgaagatccgcgcgt 3211
 DB 76 GTCACGCTCCACCTGCCACCTCCGACCGCTCTCAGGCTACCGACAGATTCGGGCGT 135
 QY 3212 ccaggttttcaaggccttgagcccccgaacccgcgcgtgacgtgcgcgcacatgagcgc 3271
 DB 136 CCAGGTTTCAAGGCTGAGGCCCGGAAACCGGGGATGCGGCTCTGCGATGAGACT 195
 QY 3272 gtccgaagcagc 3331
 DB 196 GTGGAGAGCGAGCGCGCCCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 255
 QY 3332 ctctccatggtgtgtgacgc 3391
 DB 256 CTTCTCCAGT 315
 QY 3392 gtagatggcccttgagcaagaactgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3451
 DB 316 GAGATGGCCCTGAGCAAGACTGCGCGCTCGCGGACGCTCGGGGGGCTCCCGGCC 375
 QY 3452 gccatggcgttaagctgtgc 3486
 DB 376 GCCATGGCGCCCAAGCGTCCGCTCATGGTCAT 410

RESULT 15
 AF113844 2028 bp mRNA linear PLN 20-APR-1999
 LOCUS Triticum aestivum granule-bound starch synthase precursor (Wx-D1)
 DEFINITION mRNA, Wx-D1b allele, complete cds.
 ACCESSION AF113844
 VERSION AF113844.1 GI:4588608
 KEYWORDS
 SOURCE bread wheat.
 ORGANISM Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Poideae; Triticeae; Triticum.
 1 (bases 1 to 2028)
 TITLE Vrinten, P., Nakamura, T. and Yamamori, M.
 JOURNAL Molecular characterization of waxy mutations in wheat
 MEDLINE Mol. Gen. Genet. 261 (3), 463-471 (1999)
 99254805
 2 (bases 1 to 2028)
 Vrinten, P.L. and Nakamura, T.
 AUTHORS Direct Submission
 TITLE Submitted (14-DEC-1998) Crop Breeding, Tohoku National Experiment
 JOURNAL Station, Shimo-Kuriyagawa, Morioka, Iwate 020-01, Japan
 location/Qualifiers
 1..2028
 /organism="Triticum aestivum"
 /cultivar="Waxy"
 /db_xref="taxon:4565"
 1..2028
 /gene="Wx-D1"
 /note="Wx-D1b null allele of waxy wheat; spontaneously
 mutated Wx-D1 gene"

CDS
 /allele="Wx-D1b"
 30..1754
 /gene="Wx-D1"
 /note="Waxy protein; non functional"
 /codon_start=1
 /product="granule-bound starch synthase precursor"
 /protein_id="A026156.1"
 /db_xref="GI:4588609"

BASE COUNT 448 a 614 c 628 g 338 t
 ORIGIN

Query Match 8.3%; Score 315.2; DB 8; Length 2028;
 Best Local Similarity 89.9%; Pred. No. 6.4e-50;
 Matches 338; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 3111 gctcgaagtgcacacacccctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3170
 DB 1 GCTGCAAGTGTAGCCACACCCCTGCGCGCCATGCGGCTCTGTCAGCTCCAGCTCGCCA 60
 QY 3171 cctccgcacacgtctcaggttcacgaagatccgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3230
 DB 61 CTTCTCCAGT 120
 QY 3231 ggc 3290
 DB 121 GGCCTGGAGCCCGCGGAGTGGCGCTCTCGGATGAGACGCTCGGACCTAGCGCCGCC 180
 QY 3291 caaagcaagcagaagaacccgc 3350
 DB 181 CAACGCAAGCCCGGAAAGCGCACCCGCGGAGCGCGGTCTCTCAGTGTGAGCGCG 240
 QY 3351 ccacgc 3410
 DB 241 CCACGCGGAGCGCGGCGCATGAACTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 300
 QY 3411 agactgacgc 3470
 DB 301 AGACCGCGCGCTCGCGGAGCTCTCGGGGCGCTCCCGCCACGACATGCGCGCAACGGCGC 360
 QY 3471 gccacgc 3486
 DB 361 ACCGGGTCTATGGTCAT 376

Search completed: July 31, 2002, 14:00:36
 Job time: 17478 sec

Thu Aug 1 08:29:55 2002

us-09-899-718a-1.rge

Page 16